PCT/US 97/03313

# U.S. PATENT APPLICATION

60/013,612
PROVISIONAL

02/28/96

MARIE A. FACELLA, RALEIGH, NC; SANDRA L. VOLRATH, DURHAM, NC; ERIC R. WARD, DURHAM, NC.

REC'D 29 APR 1997
WIPO PCT

**CONTINUING DATA*************************
VERIFIED

**FOREIGN/PCT APPLICATIONS************
VERIFIED

PRIORITY DOCUMENT

FOREIGN FILING LICENSE GRANTED 04/23/96

| STATE OR COUNTRY | SHEETS DRAWINGS | TOTAL CLAIMS | INDEPENDENT CLAIMS | FILING FEE RECEIVED | ATTORNEY DOCKET NO. |
|---|---|---|---|---|---|
| NC | 0 | | | $150.00 | PROV/CGC1846 |

MICHAEL W GLYNN
CIBA-GEIGY CORPORATION
520 WHITE PLAINS ROAD
P O BOX 2005
TARRYTOWN NY 10591-9005

PROMOTERS FROM PLANT PROTOPORPHYRINOGEN OXIDASE GENES

This is to certify that annexed hereto is a true copy from the records of the United States Patent and Trademark Office of the application which is identified above.

By authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

MAR 17 1997

Date Certifying Officer

PATENT APPLICATION SERIAL NO. 08/010512

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE
FEE RECORD SHEET

270 SB 07-0590 04/17/96 08013A12
29053 114 150.00CH PROV/CEC1046

PTO-1556
(5/87)

701251[redacted]

MAIL ROOM FEB 28 1996 U.S. PAT. & TRADEMARK OFF.

| Docket Number | PROV/CGC1846 |
|---|---|
| CERTIFICATE OF MAILING BY "EXPRESS MAIL" | |
| EG826034388US "Express Mail" mailing number | February 28, 1996 Date of Deposit |
| I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Commissioner of Patents and Trademarks, Washington, D.C. 20231. | |
| Priscilla Tarleton (Typed or printed name of person mailing application) | Priscilla Tarleton (Signature of person mailing application) |

To: Box: Provisional Patent Applications
Assistant Commissioner of Patents
Washington, D.C. 20231

# PROVISIONAL PATENT APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION for patent under 37 CFR 1.53 (b)(2).

INVENTOR(S)/APPLICANT(S)

| | LAST NAME | FIRST NAME | MIDDLE INITIAL | RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY) |
|---|---|---|---|---|
| 1-00 | Pacella | Maria | A. | Raleigh, North Carolina |
| 2-00 | Volrath | Sandra | L. | Durham, North Carolina |
| 3-00 | Ward | Eric | R. | Durham, North Carolina |

>NC

TITLE OF THE INVENTION (280 characters max)

Promoters from Plant Protoporphyrinogen Oxidase Genes

CORRESPONDENCE ADDRESS

Michael W. Glynn
Ciba-Geigy Corporation
Patent Department

| Mailing Address | State | Zip Code | Country |
|---|---|---|---|
| 520 White Plains Road P.O. Box 2005 Tarrytown | NEW YORK | 10591-9005 | U.S.A. |

ENCLOSED APPLICATION PARTS (check all that apply)

[X] 53 pages of Specification (and any claims)　　[X] 1 pages of Abstract (page 53)

[ ] ___ sheets of Drawing(s)　　[ ] Other (specify)

METHOD OF PAYMENT

| The Commissioner is hereby authorized to charge filing fees and additional fees required to Deposit Account Number: 07-0580. | PROVISIONAL FILING FEE AMOUNT: ($)150.00 |
|---|---|

[ ] U.S. Government agency and contract number: ________________ (if the invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.)

Respectfully submitted,

James Scott Elmer

NAME: James Scott Elmer
ATTORNEY/AGENT
REGISTRATION NO. 36,129

Tel No. (919) 541-8614
Date: February 28, 1996

[ ] Additional inventors are being named on separately numbered sheets attached hereto.

103 11/95

9/013512 [redacted]

# PROMOTERS FROM PLANT PROTOPORPHYRINOGEN OXIDASE GENES

## FIELD OF THE INVENTION

This invention relates to novel DNA sequences which function as promoters of transcription of associated DNA sequences in plants. More specifically, this invention relates to novel promoters which are naturally associated with plant protoporphyrinogen oxidase (protox) coding sequences.

## BACKGROUND OF THE INVENTION

### I. The Protox Enzyme and its Involvement in the Chlorophyll/Heme Biosynthetic Pathway

The biosynthetic pathways which lead to the production of chlorophyll and heme share a number of common steps. Chlorophyll is a light harvesting pigment present in all green photosynthetic organisms. Heme is a cofactor of hemoglobin, cytochromes, P450 mixed-function oxygenases, peroxidases, and catalases (*see, e.g.* Lehninger, <u>Biochemistry</u>. Worth Publishers, New York (1975)), and is therefore a necessary component for all aerobic organisms.

The last common step in chlorophyll and heme biosynthesis is the oxidation of protoporphyrinogen IX to protoporphyrin IX. Protoporphyrinogen oxidase (referred to herein as "protox") is the enzyme which catalyzes this last oxidation step (Matringe *et al., Biochem. J. 260:* 231 (1989)).

The protox enzyme has been purified either partially or completely from a number of organisms including the yeast *Saccharomyces cerevisiae* (Labbe-Bois and Labbe, In <u>Biosynthesis of Heme and Chlorophyll</u>, E.H. Dailey, ed. McGraw Hill: New York, pp. 235-285 (1990)), barley etioplasts (Jacobs and Jacobs, *Biochem. J. 244:* 219 (1987)), and mouse liver (Dailey and Karr, *Biochem. 26*: 2697 (1987)). Genes encoding protox have been isolated from two prokaryotic organisms, *Escherichia coli* (Sasarman *et al., Can. J. Microbiol. 39:* 1155 (1993)) and *Bacillus subtilis* (Dailey *et al., J. Biol. Chem. 269:* 813 (1994)). These genes share no sequence similarity; neither do their predicted protein products share any amino acid sequence identity. The *E. coli*

protein is approximately 21 kDa, and associates with the cell membrane. The *B. subtilis* protein is 51 kDa, and is a soluble, cytoplasmic activity.

Protox encoding cDNAs have now also been isolated from humans (*see* Nishimura *et al.*, *J. Biol. Chem. 270(14)*: 8076-8080 (1995) and plants (International application no. PCT/IB95/00452 filed June 8, 1995, published Dec. 21, 1995 as WO 95/34659).

## II. The Protox Gene as a Herbicide Target

The use of herbicides to control undesirable vegetation such as weeds or plants in crops has become almost a universal practice. The relevant market exceeds a billion dollars annually. Despite this extensive use, weed control remains a significant and costly problem for farmers.

Effective use of herbicides requires sound management. For instance, time and method of application and stage of weed plant development are critical to getting good weed control with herbicides. Since various weed species are resistant to herbicides, the production of effective herbicides becomes increasingly important.

Unfortunately, herbicides that exhibit greater potency, broader weed spectrum and more rapid degradation in soil can also have greater crop phytotoxicity. One solution applied to this problem has been to develop crops which are resistant or tolerant to herbicides. Crop hybrids or varieties resistant to the herbicides allow for the use of the herbicides without attendant risk of damage to the crop. Development of resistance can allow application of a herbicide to a crop where its use was previously precluded or limited (*e.g.* to pre-emergence use) due to sensitivity of the crop to the herbicide. For example, U.S. Patent No. 4,761,373 to Anderson *et al.* is directed to plants resistant to various imidazolinone or sulfonamide herbicides. The resistance is conferred by an altered acetohydroxyacid synthase (AHAS) enzyme. U.S. Patent No. 4,975,374 to Goodman *et al.* relates to plant cells and plants containing a gene encoding a mutant glutamine synthetase (GS) resistant to inhibition by herbicides that were known to inhibit GS, *e.g.* phosphinothricin and methionine sulfoximine. U.S. Patent No. 5,013,659 to Bedbrook *et al.* is directed to plants that express a mutant acetolactate synthase which renders the plants resistant to inhibition by sulfonylurea herbicides. U.S. Patent No. 5,162,602 to Somers *et al.* discloses plants

tolerant to inhibition by cyclohexanediones and aryloxyphenoxypropanoic acid herbicides. The tolerance is conferred by an altered acetyl coenzyme A carboxylase(ACCase).

The protox enzyme serves as the target for a variety of herbicidal compounds. The herbicides that inhibit protox include many different structural classes of molecules (Duke *et al.*, *Weed Sci.* *39*: 465 (1991); Nandihalli *et al.*, *Pesticide Biochem. Physiol.* *43:* 193 (1992); Matringe *et al.*, *FEBS Lett.* *245:* 35 (1989); Yanase and Andoh, *Pesticide Biochem. Physiol.* *35*: 70 (1989)). These herbicidal compounds include the diphenylethers {e.g. acifluorfen, 5-[2-chloro-4-(trifluoromethyl)phenoxy]-2-nitrobenzoic acid; its methyl ester; or oxyfluorfen, 2-chloro-1-(3-ethoxy-4-nitrophenoxy)-4-(trifluorobenzene)}, oxidiazoles, (e.g. oxidiazon, 3-[2,4-dichloro-5-(1-methylethoxy)phenyl]-5-(1,1-dimethylethyl)-1,3,4-oxadiazol-2-(3*H*)-one), cyclic imides (e.g. S-23142, *N*-(4-chloro-2-fluoro-5-propargyloxyphenyl)-3,4,5,6-tetrahydrophthalimide; chlorophthalim, *N*-(4-chlorophenyl)-3,4,5,6-tetrahydrophthalimide), phenyl pyrazoles (e.g. TNPP-ethyl, ethyl 2-[1-(2,3,4-trichlorophenyl)-4-nitropyrazolyl-5-oxy]propionate; M&B 39279), pyridine derivatives (e.g. LS 82-556), and phenopylate and its *O*-phenylpyrrolidino- and piperidinocarbamate analogs. Many of these compounds competitively inhibit the normal reaction catalyzed by the enzyme, apparently acting as substrate analogs.

Typically, the inhibitory effect on protox is determined by measuring fluorescence at about 622 to 635 nM, after excitation at about 395 to 410 nM (see, *e.g.* Jacobs and Jacobs, *Enzyme 28:* 206 (1982); Sherman *et al.*, *Plant Physiol.* *97:* 280 (1991)). This assay is based on the fact that protoporphyrin IX is a fluorescent pigment, and protoporphyrinogen IX is nonfluorescent.

The predicted mode of action of protox-inhibiting herbicides involves the accumulation of protoporphyrinogen IX in the chloroplast. This accumulation is thought to lead to leakage of protoporphyrinogen IX into the cytosol where it is oxidized by a peroxidase activity to protoporphyrin IX. When exposed to light, protoporphyrin IX can cause formation of singlet oxygen in the cytosol. This singlet oxygen can in turn lead to the formation of other reactive oxygen species, which can cause lipid peroxidation and membrane disruption leading to rapid cell death (Lee *et al.*, *Plant Physiol.* *102*: 881 (1993)).

Not all protox enzymes are sensitive to herbicides which inhibit plant protox enzymes. Both of the protox enzymes encoded by genes isolated from *Escherichia coli* (Sasarman *et al.*,

*Can. J. Microbiol.* *39:* 1155 (1993)) and *Bacillus subtilis* (Dailey *et al.*, *J. Biol. Chem.* *269:* 813 (1994)) are resistant to these herbicidal inhibitors. In addition, mutants of the unicellular alga *Chlamydomonas reinhardtii* resistant to the phenylimide herbicide S-23142 have been reported (Kataoka *et al.*, *J. Pesticide Sci.* *15:* 449 (1990); Shibata *et al.*, In Research in Photosynthesis, Vol. III, N. Murata, ed. Kluwer:Netherlands. pp. 567-570 (1992)). At least one of these mutants appears to have an altered protox activity that is resistant not only to the herbicidal inhibitor on which the mutant was selected, but also to other classes of protox inhibitors (Oshio *et al.*, *Z. Naturforsch.* *48c:* 339 (1993); Sato *et al.*, In ACS Symposium on Porphyric Pesticides, S. Duke, ed. ACS Press: Washington, D.C. (1994)). A mutant tobacco cell line has also been reported that is resistant to the inhibitor S-21432 (Che *et al.*, *Z. Naturforsch.* *48c:* 350 (1993). In addition, modified, inhibitor-resistant forms of plant protox coding sequences have been described in international application no. PCT/IB95/00452 filed June 8, 1995, published Dec. 21, 1995 as WO 95/34659.

### III. Regulation of Protox Gene Expression

The bulk of the research related to the protox gene which has been conducted thus far has focused upon the coding sequence and modifications to this enzyme which may render it resistant to protox inhibitors. No information is available in the art with regard to the regulatory elements which control and promote the expression of protox coding sequences in plants.

## SUMMARY OF THE INVENTION

The present invention is based on the discovery that the promoter regions naturally associated with the plant protoporphyrinogen oxidase (protox) coding sequences, referred to herein generally as the "protox promoter", are useful for promoting expression of a heterologous coding sequence in a plant.

In accordance with this discovery, the present invention provides an isolated DNA molecule comprising a plant protox promoter. The present invention further provides a chimeric

gene comprising a plant protox promoter operably linked to a heterologous coding sequence. Plant tissue and plants containing such a chimeric gene are also provided.

In one aspect of the invention the protox promoter is used to express herbicide resistant forms of herbicide target proteins in a plant to confer tolerance to the herbicide. According to this aspect, the protox promoter may be operably linkrd to a coding sequence for a herbicide-resistant plant protox protein which is resistant to inhibitors of unmodified plant protox protein.

## DESCRIPTION OF THE SEQUENCE LISTING

| | |
|---|---|
| SEQ ID No. 1: | DNA coding sequence for an *Arabidopsis thaliana* protox-1 protein. |
| SEQ ID No. 2: | *Arabidopsis thaliana* protox-1 amino acid sequence encoded by SEQ ID No. 1. |
| SEQ ID No. 3: | DNA coding sequence for an *Arabidopsis thaliana* protox-2 protein. |
| SEQ ID No. 4: | *Arabidopsis thaliana* protox-2 amino acid sequence encoded by SEQ ID No.3 |
| SEQ ID No. 5: | DNA coding sequence for a maize protox-1 protein. |
| SEQ ID No. 6: | Maize protox-1 amino acid sequence encoded by SEQ ID No. 5 |
| SEQ ID No. 7: | DNA coding sequence for a maize protox-2 protein. |
| SEQ ID No. 8: | Maize protox-2 amino acid sequence encoded by SEQ ID No. 7 |
| SEQ ID No. 9: | DNA coding sequence for a wheat protox-1 protein. |
| SEQ ID No. 10: | Wheat protox-1 amino acid sequence encoded by SEQ ID No. 9. |
| SEQ ID No. 11: | DNA coding sequence for a soybean protox-1 protein. |
| SEQ ID No. 12: | Soybean protox-1 protein encoded by SEQ ID No. 11. |
| SEQ ID NO. 13: | Promoter sequence from *Arabidopsis thaliana* protox-1 gene. |

## DEFINITIONS

As used herein a "plant protox promoter" is used to refer to the regulatory region which naturally occurs immediately upstream of a protoporphyrinogen oxidase (protox) coding sequence in a plant and is responsible, in its naturally occurring state, for regulating the transcription of the associated protox coding sequence. The plant protox promoter includes the DNA region directly involved in binding of RNA polymerase to initiate transcription and additional upstream regulatory cis-elements which influence the transcription of an operably linked coding sequence.

As used herein a "gene" is used to refer to a DNA molecule which includes (1) a coding sequence and (2) associated regulatory regions which promote and regulate the transcription of the coding sequence in a suitable host cell. The coding sequence may encode a useful transcript (e.g. antisense RNA) or polypeptide produced by translation of the encoded transcript. A gene includes at a minimum, in 5'-3' orientation, a promoter region, a coding sequence and a transcription terminator. A gene may also include additional regulatory regions which can occur as part of the minimal elements (e.g. leaders or signal peptides within the coding sequence) or as discrete elements (e.g. introns).

As used herein a "chimeric gene" refers to a gene which does not naturally occur wherein at least one component part is heterologous with respect to another component part. As used herein to describe the present invention a "chimeric gene" refers to a gene which includes the promoter of the invention operably linked to a heterologous coding sequence.

As used herein with reference to the relationship between a promoter and a coding sequence, the term "heterologous" is used to refer to a relationship which does not naturally occur. For instance, a coding sequence is considered heterologous with respect to a promoter sequence if it is different from the coding sequence that naturally occurs in association with the promoter sequence. This includes modified forms of coding sequences which are naturally associated with a subject promoter. Accordingly, a modified, inhibitor-resistant protox coding sequence is considered to be heterologous with respect to the promoter that is naturally associated with the unmodified, inhibitor-sensitive form of this coding sequence.

As used herein, the term "substantial sequence homology" is used to indicate that a nucleotide sequence (in the case of DNA or RNA) or an amino acid sequence (in the case of a protein or polypeptide) exhibits substantial structural and functional equivalence with another nucleotide or amino acid sequence. Any functional or structural differences between sequences having substantial sequence homology will be de minimis; that is they will not affect the ability of the sequence to function as indicated in the present application. For example, a sequence which has substantial sequence homology with a DNA sequence disclosed to be a plant protox promoter will be able to direct the same level and pattern of expression of an associated DNA sequence as the plant protox promoter. Sequences that have substantial sequence homology with the sequences disclosed herein are usually variants of the disclosed sequence, such as mutations, but may also be synthetic sequences. Structural differences are considered de minimis if there is a significant amount of sequence overlap or similarity between two or more different sequences or if the different sequences exhibit similar physical characteristics. Such characteristics can include, for example, immunological reactivity, enzyme activity, structural protein integrity, etc.

Two nucleotide sequences may have substantial sequence homology if the sequences have at least 70 percent, more preferably 80 percent and most preferably 90 percent sequence similarity between them. Two amino acid sequences have substantial sequence homology if they have at least 50 percent, preferably 70 percent, and most preferably 90 percent similarity between the active portions of the polypeptides. In the case of promoter DNA sequences, "substantial sequence homology" also refers to those fragments of a promoter DNA sequence that are able to operate to promote the expression of associated DNA sequences. Such operable fragments of a promoter DNA sequence may be derived from the promoter DNA sequence, for example, by cleaving the promoter DNA sequence using restriction enzymes, synthesizing in accordance with the sequence of the promoter DNA sequence, or may be obtained through the use of PCR technology. Mullis et al., Meth. Enzymol., 155:335-350 (1987); Erlich (ed.), PCR Technology, Stockton Press (New York 1989).

A promoter DNA sequence is said to be "operably linked" to a second DNA sequence if the two are situated such that the promoter DNA sequence influences the transcription or translation of the second DNA sequence. For example, if the second DNA sequence codes for

the production of a protein, the promoter DNA sequence would be operably linked to the second DNA sequence if the promoter DNA sequence affects the expression of the protein product from the second DNA sequence. For example, in a DNA sequence comprising a promoter DNA sequence physically attached to a coding DNA sequence in the same chimeric construct, the two sequences are likely to be operably linked.

## DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to promoter DNA sequences which are naturally associated with coding sequences for plant protoporphyrinogen oxidase (referred to herein as "protox"; *see* international application no. PCT/IB95/00452 filed June 8, 1995, published Dec. 21, 1995 as WO 95/34659 and co-pending provisional application entitled " DNA Molecules Encoding Plant Protoporphyrinogen Oxidase and Inhibitor Resistant Mutants Thereof" filed on the same day as the instant application). These protox promoter sequences have been found to be useful for the expression of a heterologous coding sequence in a plant.

The promoter sequence for the *Arabidopsis thaliana* protox-1 coding sequence (SEQ ID No. 1) is provided as SEQ ID No. 13. Isolation of this promoter from a genomic library using the associated coding sequence as a probe is described in Example 1. This same approach can be used to isolate the promoter sequence from any plant protox gene. Any protox coding sequence which shares sufficient homology to hybridize to the protox coding sequence associated with the promoter of interest may be used as a probe in this approach. Since the respective protox-1 and protox-2 coding sequences from all plants are contemplated to share this requisite degree of homology, the choice of which protox coding sequence is used as a probe is not considered critical. However, for optimal hybridization results it is preferable to use the most closely related protox coding sequence. Most preferably, the coding sequence used as a probe is from the same plant species as the protox promoter of interest and is the coding sequence naturally associated with the promoter.

The plant protox promoter of the present invention includes the *Arabidopsis* protox-1 promoter sequence set forth in SEQ Id No. 13 as well as corresponding protox-1 promoter sequences available from other plant species as indicated above. The present invention also

the production of a protein, the promoter DNA sequence would be operably linked to the second DNA sequence if the promoter DNA sequence affects the expression of the protein product from the second DNA sequence. For example, in a DNA sequence comprising a promoter DNA sequence physically attached to a coding DNA sequence in the same chimeric construct, the two sequences are likely to be operably linked.

## DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to promoter DNA sequences which are naturally associated with coding sequences for plant protoporphyrinogen oxidase (referred to herein as "protox"; *see* international application no. PCT/IB95/00452 filed June 8, 1995, published Dec. 21, 1995 as WO 95/34659 and co-pending provisional application entitled " DNA Molecules Encoding Plant Protoporphyrinogen Oxidase and Inhibitor Resistant Mutants Thereof" filed on the same day as the instant application). These protox promoter sequences have been found to be useful for the expression of a heterologous coding sequence in a plant.

The promoter sequence for the *Arabidopsis thaliana* protox-1 coding sequence (SEQ ID No. 1) is provided as SEQ ID No. 13. Isolation of this promoter from a genomic library using the associated coding sequence as a probe is described in Example 1. This same approach can be used to isolate the promoter sequence from any plant protox gene. Any protox coding sequence which shares sufficient homology to hybridize to the protox coding sequence associated with the promoter of interest may be used as a probe in this approach. Since the respective protox-1 and protox-2 coding sequences from all plants are contemplated to share this requisite degree of homology, the choice of which protox coding sequence is used as a probe is not considered critical. However, for optimal hybridization results it is preferable to use the most closely related protox coding sequence. Most preferably, the coding sequence used as a probe is from the same plant species as the protox promoter of interest and is the coding sequence naturally associated with the promoter.

The plant protox promoter of the present invention includes the *Arabidopsis* protox-1 promoter sequence set forth in SEQ Id No. 13 as well as corresponding protox-1 promoter sequences available from other plant species as indicated above. The present invention also

includes functional fragments of these DNA sequences which retain the ability to regulate expression of an operably linked coding sequence in the same manner as the exemplified protox promoter sequence. Such functional fragments may be identified through deletion analyses or other standard techniques used in the art to identify protox promoter activity (*see, e.g.* pages 546-549 of "Genes IV", ed. by Lewin, Oxford Univ. Press (1990)). The present invention also includes DNA sequences having substantial sequence homology with the protox promoters available from plant genes which confer an equivalent level and pattern of expression upon an operably linked sequence. Such promoter sequences may be obtained through modification of the protox promoters isolated from plant genes and are considered functionally equivalent derivatives of the plant protox promoters.

As illustrated in the examples below, the DNA sequences, vectors and transgenic plants of the present invention comprise a promoter sequence derived from a plant protox gene. The protox promoter DNA sequences are preferably linked operably to a coding DNA sequence, for example a DNA sequence which is transcribed into a useful RNA transcript such as an antisense transcript, or a coding sequence which is ultimately expressed in the production of a useful protein product.

In a preferred embodiment, the protox promoter is used to direct the expression of a modified herbicide target enzyme which is resistant to herbicides at levels that inhibit the corresponding unmodified version of the enzyme. Such modified herbicide-resistant enzymes include herbicide-resistant forms of imidazoleglycerol phosphate dehyratase (IGPD; *see* WO 9426909 published Nov. 24, 1994), EPSP synthase (*see* U.S. Pat. Nos. 4,535,060; 4,769,061; 4,940,835 and EP 550,633), glutamine synthetase (GS; *see* U.S. Patent No. 4,975,374), acetyl coenzyme A carboxylase(ACCase; *see* U.S. Patent No. 5,162,602), and acetolactate synthase (*see* U.S. Patent Nos. 4,761,373; 5,304,732; 5,331,107; 5,013,659; 5,141,870; and 5,378,824). In a most preferred embodiment, the protox promoter is used to direct the expression of a modified protox enzyme which is resistant to protox inhibitors as illustrated in Examples 2-3 (*see also* International application no. PCT/IB95/00452 filed June 8, 1995, published Dec. 21, 1995 as WO 95/34659 whose relevant parts are herein incorporated by reference; *see also* co-pending

application entitled " DNA Molecules Encoding Plant Protoporphyrinogen Oxidase and Inhibitor Resistant Mutants Thereof" filed on the same day as the instant application).

The transgenic plants of the present invention may be transformed by any method of transformation known in the art. These methods include, for instance, transformation by direct infection or co-cultivation of plants, plant tissue or cells with *Agrobacterium tumefaciens*; Horsch *et al., Science, 225:* 1229 (1985); Marton, "Cell Culture and Somatic Cell Genetic of Plants", vol 1, pp 514-521 (1984); direct gene transfer into protoplasts; Paszkowski *et al., EMBO J. 12:* 2717 (1984); Loerz *et al., Mol. Gen. & Genet. 1199*:178 (1985); Fromm *et al., Nature 319*:719 (1986); microprojectile bombardment, Klein *et al., Bio/Technology, 6*:559-563 (1988); injection into protoplasts cultured cells and tissues, Reich *et al., Bio/Technology, 4*:1001-1004 (1986); or injection into meristematic tissues of seedlings and plants as described by De La Pena *et al., Nature, 325*:274-276 (1987); Hooykaas-Van Slogteren *et al., Nature, 311*:763-764 (1984); Grimsley *et al., Bio/Technology, 6*:185 (1988); and Grimsley *et al., Nature, 325*:177 (1988).

The invention is illustrated in more detail by the following examples, without implying any restriction to what is described therein.

## EXAMPLES

**EXAMPLE 1: Isolation of the Arabidopsis thaliana Protox-1 promoter sequence**

A Lambda Zap II genomic DNA library prepared from *Arabidopsis thaliana* (Columbia, whole plant) was purchased from Stratagene. Approximately 125,000 phage were plated at a density of 25,000 pfu (plaque forming units) per 15 cm Petri dish and duplicate lifts were made onto Colony/Plaque Screen membranes (NEN Dupont). The plaque lifts were probed with the Arabidopsis Protox-1 cDNA (SEQ ID No. 1) labeled with 32P-dCTP by the random priming method (Life Technologies). Hybridization and wash conditions were at 65° C as described in Church and Gilbert, *Proc. Natl. Acad. Sci. USA 81:* 1991-1995 (1984). Positively hybridizing plaques were purified and *in vivo* excised into pBluescript plasmids. Sequence from the genomic DNA inserts was determined by the chain termination method using dideoxy terminators labeled with fluorescent dyes (Applied Biosystems, Inc.). One clone, AraPT1Pro, was determined to contain 580 bp of Arabidopsis sequence upstream from the initiating methionine (ATG) of the

Protox-1 protein coding sequence. This clone also contains coding sequence and introns that extend to bp 1241 of the Protox-1 cDNA sequence. The 580 bp 5' noncoding fragment is the putative Arabidopsis Protox-1 promoter, and the sequence is set forth in SEQ ID No. 13.

AraPT1Pro was deposited December 14, 1995, as pWDC-11 (NRRL #B-21515)

**EXAMPLE 2: Construction of plant transformation vectors expressing altered Protox-1 genes behind the native Arabidopsis Protox-1 promoter**

A full-length cDNA of the appropriate altered *Arabidopsis* Protox-1 cDNA is isolated as an EcoRI-XhoI partial digest fragment and cloned into the plant expression vector pCGN1761ENX (see Example 9 of International application no. PCT/IB95/00452 filed June 8, 1995, published Dec. 21, 1995 as WO 95/34659). This plasmid is digested with NcoI and BamHI to produce a fragment comprised of the complete Protox-1 cDNA plus a transcription terminator from the 3' untranslated sequence of the tml gene of *Agrobacterium tumefaciens*. The AraPT1Pro plasmid described above is digested with NcoI and BamHI to produce a fragment comprised of pBluescript and the 580 bp putative Arabidopsis Protox-1 promoter. Ligation of these two fragments produces a fusion of the altered protox cDNA to the native protox promoter. The expression cassette containing the Protox-1 promoter/Protox-1 cDNA/tml terminator fusion is excised by digestion with KpnI and cloned into the binary vector pCIB200. The binary plasmid is transformed by electroporation into Agrobacterium and then into Arabidopsis using the vacuum infiltration method (Bechtold *et al. C.R. Acad. Sci. Paris 316:* 1194-1199 (1993)). Transformants expressing altered protox genes are selected on kanamycin or on various concentrations of protox inhibiting herbicide.

**EXAMPLE 3: Production of herbicide tolerant plants by expression of a native Protox-1 promoter/altered Protox-1 fusion**

Using the procedure described above, an *Arabidopsis* Protox-1 cDNA containing a TAC to ATG (Tyrosine to Methionine) change at nucleotides 1306-1308 in the Protox-1 sequence (SEQ ID No.1) was fused to the native Protox-1 promoter fragment and transformed into *Arabidopsis thaliana*. This altered Protox-1 enzyme (AraC-2Met) has been shown to be >10fold

more tolerant to various protox-inhibiting herbicides than the naturally occurring enzyme when tested in a bacterial expression system (*see* Example 5 of copending U.S. application entitled " DNA Molecules Encoding Plant Protoporphyrinogen Oxidase and Inhibitor Resistant Mutants Thereof" filed on the same day as the instant application). Seed from the vacuum infiltrated plants was collected and plated on a range (10.0nM-1.0uM) of a protox inhibitory aryluracil herbicide of formula XVII. Multiple experiments with wild type *Arabidopsis* have shown that a 10.0nM concentration of this compound is sufficient to prevent normal seedling germination. Transgenic seeds expressing the AraC-2Met altered enzyme fused to the native Protox-1 promoter produced normal Arabidopsis seedlings at herbicide concentrations up to 500nM, indicating at least 50-fold higher herbicide tolerance when compared to wild-type Arabidopsis. This promoter/altered protox enzyme fusion therefore functions as an effective selectable marker for plant transformation. Several of the plants that germinated on 100.0nM of protox-inhibiting herbicide were transplanted to soil, grown 2-3 weeks, and tested in a spray assay with various concentrations of the protox-inhibiting herbicide. When compared to empty vector control transformants, the AraPT1Pro/AraC-2Met transgenics were >10fold more tolerant to the herbicide spray.

EXAMPLE 4: Isolation of Maize Protox-1 promoter sequences

A *Zea mays* (Missouri 17 inbred, etiolated seedlings) genomic DNA library in the Lambda FIX II vector was purchased from Stratagene. Approximately 250,000 pfu of the library was plated at a density of 50,000 phage per 15 cm plate and duplicate lifts were made onto Colony/Plaque screen membranes (NEN Dupont).The plaque lifts were probed with the maize Protox-1 cDNA (SEQ ID No. 5) labeled with 32P-dCTP by the random priming method (Life Technologies). Hybridization and wash conditions were at 65° C as described in Church and Gilbert, *Proc. Natl. Acad. Sci. USA 81:* 1991-1995 (1984). Positively hybridizing plaques were purified and rescreened with a 210 bp EcoRI-NcoI fragment from the 5' end of the maize Protox-1 cDNA. Lambda phage DNA was isolated from three phage that hybridized to the 5' fragment using the Wizard Lambda Preps DNA Purification System (Promega). Restriction analysis and hybridization to the 5' maize fragment indicated that two of the phage clones are derived from the

same gene, while a third may represent a second maize Protox-1 gene. Hybridizing fragments from both types of phage are subcloned into a pBluescript vector for sequence analysis.

## EXAMPLE 5: Construction of Plant Transformation Vectors

Numerous transformation vectors are available for plant transformation, and the promoters and chimeric genes of this invention can be used in conjunction with any such vectors. The selection of vector for use will depend upon the preferred transformation technique and the target species for transformation. For certain target species, different antibiotic or herbicide selection markers may be preferred. Selection markers used routinely in transformation include the *nptII* gene which confers resistance to kanamycin and related antibiotics (Messing & Vierra, *Gene 19:* 259-268 (1982); Bevan *et al.*, *Nature 304:*184-187 (1983)), the *bar* gene which confers resistance to the herbicide phosphinothricin (White *et al.*, *Nucl Acids Res 18:* 1062 (1990), Spencer *et al.* Theor Appl Genet 79: 625-631(1990)), the *hph* gene which confers resistance to the antibiotic hygromycin (Blochinger & Diggelmann, *Mol Cell Biol 4:* 2929-2931), and the *dhfr* gene, which confers resistance to methotrexate (Bourouis *et al.*, *EMBO J. 2(7):* 1099-1104 (1983)).

### (1) Construction of Vectors Suitable for *Agrobacterium* Transformation

Many vectors are available for transformation using *Agrobacterium tumefaciens*. These typically carry at least one T-DNA border sequence and include vectors such as pBIN19 (Bevan, *Nucl. Acids Res.* (1984)) and pXYZ. Below the construction of two typical vectors is described.

<u>Construction of pCIB200 and pCIB2001</u>

The binary vectors pCIB200 and pCIB2001 are used for the construction of recombinant vectors for use with *Agrobacterium* and was constructed in the following manner. pTJS75kan was created by *NarI* digestion of pTJS75 (Schmidhauser & Helinski, *J Bacteriol. 164:* 446-455

(1985)) allowing excision of the tetracycline-resistance gene, followed by insertion of an *AccI* fragment from pUC4K carrying an NPTII (Messing & Vierra, *Gene 19*: 259-268 (1982); Bevan *et al.*, *Nature 304*: 184-187 (1983); McBride *et al.*, *Plant Molecular Biology 14*: 266-276 (1990)). *XhoI* linkers were ligated to the *EcoRV* fragment of pCIB7 which contains the left and right T-DNA borders, a plant selectable *nos/nptII* chimeric gene and the pUC polylinker (Rothstein *et al.*, *Gene 53:* 153-161 (1987)), and the *XhoI*-digested fragment was cloned into *SalI*-digested pTJS75kan to create pCIB200 (see also EP 0 332 104, example 19 [1338]). pCIB200 contains the following unique polylinker restriction sites: *EcoRI, SstI, KpnI, BglII, XbaI,* and *SalI*. *pCIB2001* is a derivative of pCIB200 which created by the insertion into the polylinker of additional restriction sites. Unique restriction sites in the polylinker of pCIB2001 are *EcoRI, SstI, KpnI, BglII, XbaI, SalI, MluI, BclI, AvrII, ApaI, HpaI,* and *StuI*. pCIB2001, in addition to containing these unique restriction sites also has plant and bacterial kanamycin selection, left and right T-DNA borders for *Agrobacterium*-mediated transformation, the RK2-derived *trfA* function for mobilization between *E. coli* and other hosts, and the *OriT* and *OriV* functions also from RK2. The pCIB2001 polylinker is suitable for the cloning of plant expression cassettes containing their own regulatory signals.

Construction of pCIB10 and Hygromycin Selection Derivatives thereof

The binary vector pCIB10 contains a gene encoding kanamycin resistance for selection in plants, T-DNA right and left border sequences and incorporates sequences from the wide host-range plasmid pRK252 allowing it to replicate in both *E. coli* and *Agrobacterium*. Its construction is described by Rothstein *et al.*, *Gene 53*: 153-161 (1987). Various derivatives of pCIB10 have been constructed which incorporate the gene for hygromycin B phosphotransferase described by Gritz *et al.*, *Gene 25:* 179-188 (1983)). These derivatives enable selection of transgenic plant cells on hygromycin only (pCIB743), or hygromycin and kanamycin (pCIB715, pCIB717).

(2) **Construction of Vectors Suitable for non-*Agrobacterium* Transformation.**

Transformation without the use of *Agrobacterium tumefaciens* circumvents the requirement for T-DNA sequences in the chosen transformation vector and consequently vectors lacking these sequences can be utilized in addition to vectors such as the ones described above which contain T-DNA sequences. Transformation techniques which do not rely on *Agrobacterium* include transformation via particle bombardment, protoplast uptake (*e.g.* PEG and electroporation) and microinjection. The choice of vector depends largely on the preferred selection for the species being transformed. Below, the construction of some typical vectors is described.

Construction of pCIB3064

pCIB3064 is a pUC-derived vector suitable for direct gene transfer techniques in combination with selection by the herbicide basta (or phosphinothricin). The plasmid pCIB246 comprises the CaMV 35S promoter in operational fusion to the *E. coli* GUS gene and the CaMV 35S transcriptional terminator and is described in the PCT published application WO 93/07278. The 35S promoter of this vector contains two ATG sequences 5' of the start site. These sites were mutated using standard PCR techniques in such a way as to remove the ATGs and generate the restriction sites *SspI* and *PvuII*. The new restriction sites were 96 and 37 bp away from the unique *SalI* site and 101 and 42 bp away from the actual start site. The resultant derivative of pCIB246 was designated pCIB3025. The GUS gene was then excised from pCIB3025 by digestion with *SalI* and *SacI*, the termini rendered blunt and religated to generate plasmid pCIB3060. The plasmid pJIT82 was obtained from the John Innes Centre, Norwich and the 400 bp *SmaI* fragment containing the *bar* gene from *Streptomyces viridochromogenes* was excised and inserted into the *HpaI* site of pCIB3060 (Thompson *et al.* EMBO J 6: 2519-2523 (1987)). This generated pCIB3064 which comprises the *bar* gene under the control or the CaMV 35S promoter and terminator for herbicide selection, a gene fro ampicillin resistance (for selection in

*E. coli*) and a polylinker with the unique sites *SphI, PstI, HindIII,* and *BamHI*. This vector is suitable for the cloning of plant expression cassettes containing their own regulatory signals.

Construction of pSOG19 and pSOG35

pSOG35 is a transformation vector which utilizes the *E. coli* gene dihydrofolate reductase (DHFR) as a selectable marker conferring resistance to methotrexate. PCR was used to amplify the 35S promoter (~800 bp), intron 6 from the maize Adh1 gene (~550 bp) and 18 bp of the GUS untranslated leader sequence from pSOG10. A 250 bp fragment encoding the *E. coli* dihydrofolate reductase type II gene was also amplified by PCR and these two PCR fragments were assembled with a *SacI-PstI* fragment from pBI221 (Clontech) which comprised the pUC19 vector backbone and the nopaline synthase terminator. Assembly of these fragments generated pSOG19 which contains the 35S promoter in fusion with the intron 6 sequence, the GUS leader, the DHFR gene and the nopaline synthase terminator. Replacement of the GUS leader in pSOG19 with the leader sequence from Maize Chlorotic Mottle Virus (MCMV) generated the vector pSOG35. pSOG19 and pSOG35 carry the pUC gene for ampicillin resistance and have *HindIII, SphI, PstI* and *EcoRI* sites available for the cloning of foreign sequences such as chimeric gene sequences containing a plant protox promoter.

EXAMPLE 12: Construction of Chimeric Genes/Plant Expression Cassettes

Coding sequences intended for expression in transgenic plants under the control of a plant protox promoter may be assembled in expression cassettes behind a suitable protox promoter and upstream of a suitable transcription terminator. The resulting chimeric genes can then be easily transferred to the plant transformation vectors described above in Example 19.

Protox Promoter Selection

In accordance with the present invention, the chimeric gene will contain a plant protox promoter. The selection of the specific protox promoter used in the chimeric gene is primarily up

to the individual researcher, although generally it will be preferable to use a protox promoter from a plant species closely related to, or most preferably identical, to the species intended to contain the resulting chimeric gene. For example, if the chimeric gene is intended to be contained in a maize plant it would be preferable to use a protox promoter from a monocotyledenous plant and most preferable to use a maize protox promoter.

Transcriptional Terminators

A variety of transcriptional terminators are available for use in expression cassettes. These are responsible for the termination of transcription beyond the transgene and its correct polyadenylation. Appropriate transcriptional terminators are those which are known to function in plants and include the CaMV 35S terminator, the *tml* terminator, the nopaline synthase terminator, the pea *rbcS* E9 terminator, as well as terminators naturally associated with the plant protox gene (i.e. "protox terminators"). These can be used in both monocotyledons and dicotyledons.

Sequences for the Enhancement or Regulation of Expression

Numerous sequences have been found to enhance gene expression from within the transcriptional unit and these sequences can be used in conjunction with the genes of this invention to increase their expression in transgenic plants.

Various intron sequences have been shown to enhance expression, particularly in monocotyledonous cells. For example, the introns of the maize *AdhI* gene have been found to significantly enhance the expression of the wild-type gene under its cognate promoter when introduced into maize cells. Intron 1 was found to be particularly effective and enhanced expression in fusion constructs with the chloramphenicol acetyltransferase gene (Callis *et al.*, Genes Develop. 1: 1183-1200 (1987)). In the same experimental system, the intron from the maize *bronze1* gene had a similar effect in enhancing expression (Callis *et al.*, *supra*). Intron sequences have been routinely incorporated into plant transformation vectors, typically within the non-translated leader.

A number of non-translated leader sequences derived from viruses are also known to enhance expression, and these are particularly effective in dicotyledonous cells. Specifically, leader sequences from Tobacco Mosaic Virus (TMV, the "W-sequence"), Maize Chlorotic Mottle Virus (MCMV), and Alfalfa Mosaic Virus (AMV) have been shown to be effective in enhancing expression (*e.g.* Gallie *et al. Nucl. Acids Res. 15*: 8693-8711 (1987); Skuzeski *et al. Plant Molec. Biol. 15*: 65-79 (1990))

<u>Targeting of the Gene Product Within the Cell</u>

Various mechanisms for targeting gene products are known to exist in plants and the sequences controlling the functioning of these mechanisms have been characterized in some detail. For example, the targeting of gene products to the chloroplast is controlled by a signal sequence found at the amino terminal end of various proteins and which is cleaved during chloroplast import yielding the mature protein (*e.g.* Comai *et al.* J. *Biol. Chem. 263*: 15104-15109 (1988)). These signal sequences can be fused to heterologous gene products to effect the import of heterologous products into the chloroplast (van den Broeck *et al. Nature 313*: 358-363 (1985)). DNA encoding for appropriate signal sequences can be isolated from the 5' end of the cDNAs encoding the RUBISCO protein, the CAB protein, the EPSP synthase enzyme, the GS2 protein and many other proteins which are known to be chloroplast localized.

Other gene products are localized to other organelles such as the mitochondrion and the peroxisome (*e.g.* Unger *et al. Plant Molec. Biol. 13*: 411-418 (1989)). The cDNAs encoding these products can also be manipulated to effect the targeting of heterologous gene products to these organelles. Examples of such sequences are the nuclear-encoded ATPases and specific aspartate amino transferase isoforms for mitochondria. Targeting to cellular protein bodies has been described by Rogers *et al., Proc. Natl. Acad. Sci. USA 82*: 6512-6516 (1985)).

In addition sequences have been characterized which cause the targeting of gene products to other cell compartments. Amino terminal sequences are responsible for targeting to the ER, the apoplast, and extracellular secretion from aleurone cells (Koehler & Ho, *Plant Cell 2*: 769-783 (1990)). Additionally, amino terminal sequences in conjunction with carboxy terminal

sequences are responsible for vacuolar targeting of gene products (Shinshi *et al., Plant Molec. Biol. 14*: 357-368 (1990)).

By the fusion of the appropriate targeting sequences described above to transgene sequences of interest it is possible to direct the transgene product to any organelle or cell compartment. For chloroplast targeting, for example, the chloroplast signal sequence from the RUBISCO gene, the CAB gene, the EPSP synthase gene, or the GS2 gene is fused in frame to the amino terminal ATG of the transgene. The signal sequence selected should include the known cleavage site and the fusion constructed should take into account any amino acids after the cleavage site which are required for cleavage. In some cases this requirement may be fulfilled by the addition of a small number of amino acids between the cleavage site and the transgene ATG or alternatively replacement of some amino acids within the transgene sequence. Fusions constructed for chloroplast import can be tested for efficacy of chloroplast uptake by *in vitro* translation of *in vitro* transcribed constructions followed by *in vitro* chloroplast uptake using techniques described by (Bartlett *et al.* In: Edelmann *et al.* (Eds.) Methods in Chloroplast Molecular Biology, Elsevier. pp 1081-1091 (1982); Wasmann *et al. Mol. Gen. Genet. 205*: 446-453 (1986)). These construction techniques are well known in the art and are equally applicable to mitochondria and peroxisomes. The choice of targeting which may be required for expression of the transgenes will depend on the cellular localization of the precursor required as the starting point for a given pathway. This will usually be cytosolic or chloroplastic, although it may is some cases be mitochondrial or peroxisomal. The products of transgene expression will not normally require targeting to the ER, the apoplast or the vacuole.

The above described mechanisms for cellular targeting can be utilized in conjunction with plant protox promoters so as to effect a specific cell targeting goal under the transcriptional regulation of a promoter which has an expression pattern different to that of the promoter from which the targeting signal derives.

**EXAMPLE 13: Transformation of Dicotyledons**

Transformation techniques for dicotyledons are well known in the art and include *Agrobacterium*-based techniques and techniques which do not require *Agrobacterium*. Non-

*Agrobacterium* techniques involve the uptake of exogenous genetic material directly by protoplasts or cells. This can be accomplished by PEG or electroporation mediated uptake, particle bombardment-mediated delivery, or microinjection. Examples of these techniques are described by Paszkowski *et al.*, *EMBO J 3:* 2717-2722 (1984), Potrykus *et al.*, *Mol. Gen. Genet. 199:* 169-177 (1985), Reich *et al.*, *Biotechnology 4:* 1001-1004 (1986), and Klein *et al.*, *Nature 327*: 70-73 (1987). In each case the transformed cells are regenerated to whole plants using standard techniques known in the art.

*Agrobacterium*-mediated transformation is a preferred technique for transformation of dicotyledons because of its high efficiency of transformation and its broad utility with many different species. The many crop species which are routinely transformable by *Agrobacterium* include tobacco, tomato, sunflower, cotton, oilseed rape, potato, soybean, alfalfa and poplar (EP 0 317 511 (cotton), EP 0 249 432 (tomato, to Calgene), WO 87/07299 (*Brassica*, to Calgene), US 4,795,855 (poplar)).

Transformation of the target plant species by recombinant *Agrobacterium* usually involves co-cultivation of the *Agrobacterium* with explants from the plant and follows protocols well known in the art. Transformed tissue is regenerated on selectable medium carrying the antibiotic or herbicide resistance marker present between the binary plasmid T-DNA borders.

**EXAMPLE 14: Transformation of Monocotyledons**

Transformation of most monocotyledon species has now also become routine. Preferred techniques include direct gene transfer into protoplasts using PEG or electroporation techniques, and particle bombardment into callus tissue. Transformations can be undertaken with a single DNA species or multiple DNA species (*i.e.* co-transformation) and both these techniques are suitable for use with this invention. Co-transformation may have the advantage of avoiding complex vector construction and of generating transgenic plants with unlinked loci for the gene of interest and the selectable marker, enabling the removal of the selectable marker in subsequent generations, should this be regarded desirable. However, a disadvantage of the use of co-transformation is the less than 100% frequency with which separate DNA species are integrated into the genome (Schocher *et al. Biotechnology 4:* 1093-1096 (1986)).

Patent Applications EP 0 292 435 (to Ciba-Geigy), EP 0 392 225 (to Ciba-Geigy), WO 93/07278 (to Ciba-Geigy) and U.S. Patent No. 5,350,689 (to Ciba-Geigy) describe techniques for the preparation of callus and protoplasts from an élite inbred line of maize, transformation of protoplasts using PEG or electroporation, and the regeneration of maize plants from transformed protoplasts. Gordon-Kamm *et al., Plant Cell 2:* 603-618 (1990)) and Fromm *et al., Biotechnology 8:* 833-839 (1990)) have published techniques for transformation of A188-derived maize line using particle bombardment. Furthermore, application WO 93/07278 (to Ciba-Geigy) and Koziel *et al., Biotechnology 11:* 194-200 (1993)) describe techniques for the transformation of élite inbred lines of maize by particle bombardment. This technique utilizes immature maize embryos of 1.5-2.5 mm length excised from a maize ear 14-15 days after pollination and a PDS-1000He Biolistics device for bombardment.

Transformation of rice can also be undertaken by direct gene transfer techniques utilizing protoplasts or particle bombardment. Protoplast-mediated transformation has been described for *Japonica*-types and *Indica*-types (Zhang *et al., Plant Cell Rep 7*: 379-384 (1988); Shimamoto *et al. Nature 338*: 274-277 (1989); Datta *et al. Biotechnology 8:* 736-740 (1990)). Both types are also routinely transformable using particle bombardment (Christou *et al. Biotechnology 9:* 957-962 (1991)).

Patent Application EP 0 332 581 (to Ciba-Geigy) describes techniques for the generation, transformation and regeneration of Pooideae protoplasts. These techniques allow the transformation of *Dactylis* and wheat. Furthermore, wheat transformation was been described by Vasil *et al., Biotechnology 10:* 667-674 (1992)) using particle bombardment into cells of type C long-term regenerable callus, and also by Vasil *et al., Biotechnology 11:* 1553-1558 (1993)) and Weeks *et al., Plant Physiol. 102*: 1077-1084 (1993) using particle bombardment of immature embryos and immature embryo-derived callus. A preferred technique for wheat transformation, however, involves the transformation of wheat by particle bombardment of immature embryos and includes either a high sucrose or a high maltose step prior to gene delivery. Prior to bombardment, any number of embryos (0.75-1 mm in length) are plated onto MS medium with 3% sucrose (Murashige & Skoog, *Physiologia Plantarum 15*: 473-497 (1962)) and 3 mg/l 2,4-D for induction of somatic embryos which is allowed to proceed in the dark. On the chosen day of

bombardment, embryos are removed from the induction medium and placed onto the osmoticum (*i.e.* induction medium with sucrose or maltose added at the desired concentration, typically 15%). The embryos are allowed to plasmolyze for 2-3 h and are then bombarded. Twenty embryos per target plate is typical, although not critical. An appropriate gene-carrying plasmid (such as pCIB3064 or pSG35) is precipitated onto micrometer size gold particles using standard procedures. Each plate of embryos is shot with the DuPont Biolistics· helium device using a burst pressure of ~1000 psi using a standard 80 mesh screen. After bombardment, the embryos are placed back into the dark to recover for about 24 h (still on osmoticum). After 24 hrs, the embryos are removed from the osmoticum and placed back onto induction medium where they stay for about a month before regeneration. Approximately one month later the embryo explants with developing embryogenic callus are transferred to regeneration medium (MS + 1 mg/liter NAA, 5 mg/liter GA), further containing the appropriate selection agent (10 mg/l basta in the case of pCIB3064 and 2 mg/l methotrexate in the case of pSOG35). After approximately one month, developed shoots are transferred to larger sterile containers known as "GA7s" which contained half-strength MS, 2% sucrose, and the same concentration of selection agent. Patent application 08/147,161 describes methods for wheat transformation and is hereby incorporated by reference.

* * * * * * * * * * * * *

While the present invention has been described with reference to specific embodiments thereof, it will be appreciated that numerous variations, modifications, and embodiments are possible, and accordingly, all such variations, modifications and embodiments are to be regarded as being within the spirit and scope of the present invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Pacalla, Marie A.
Volrath, Sandra L.
Ward, Eric R.

(ii) TITLE OF INVENTION: PROMOTERS FROM PLANT PROTOPORPHYRINOGEN OXIDASE GENES

(iii) NUMBER OF SEQUENCES: 13

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Ciba-Geigy Corporation / Patent Dept.
(B) STREET: 540 White Plains Rd.
(C) CITY: Tarrytown
(D) STATE: NY
(E) COUNTRY: USA
(F) ZIP: 10591-9005

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US TBA
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA
(A) APPLICATION NUMBER: US 08/261,198
(B) FILING DATE: 16-JUN-94

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Elmer, James Scott
(B) REGISTRATION NUMBER: 36,129
(C) REFERENCE/DOCKET NUMBER: CGC 1846/prov

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 919-541-8614
(B) TELEFAX: 919-541-8689

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1719 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 31..1644
(D) OTHER INFORMATION: /note= "Arabidopsis protox-1 cDNA; sequence from pWDC-2"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
TGACAAAATT CCGAATTCTC TGCGATTTCC ATG GAG TTA TCT CTT CTC CGT CCG      54
                                 Met Glu Leu Ser Leu Leu Arg Pro
                                  1               5

ACG ACT CAA TCG CTT CTT CCG TCG TTT TCG AAG CCC AAT CTC CGA TTA      102
Thr Thr Gln Ser Leu Leu Pro Ser Phe Ser Lys Pro Asn Leu Arg Leu
     10                  15                  20

AAT GTT TAT AAG CCT CTT AGA CTC CGT TGT TCA GTG GCC GGT GGA CCA      150
Asn Val Tyr Lys Pro Leu Arg Leu Arg Cys Ser Val Ala Gly Gly Pro
 25                  30                  35                  40

ACC GTC GGA TCT TCA AAA ATC GAA GGC GGA GGA GGC ACC ACC ATC ACG      198
Thr Val Gly Ser Ser Lys Ile Glu Gly Gly Gly Gly Thr Thr Ile Thr
                 45                  50                  55

ACG GAT TGT GTG ATT GTC GGC GGA GGT ATT AGT GGT CTT TGC ATC GCT      246
Thr Asp Cys Val Ile Val Gly Gly Gly Ile Ser Gly Leu Cys Ile Ala
             60                  65                  70

CAG GCG CTT GCT ACT AAG CAT CCT GAT GCT GCT CCG AAT TTA ATT GTG      294
Gln Ala Leu Ala Thr Lys His Pro Asp Ala Ala Pro Asn Leu Ile Val
         75                  80                  85

ACC GAG GCT AAG GAT CGT GTT GGA GGC AAC ATT ATC ACT CGT GAA GAG      342
Thr Glu Ala Lys Asp Arg Val Gly Gly Asn Ile Ile Thr Arg Glu Glu
     90                  95                 100

AAT GGT TTT CTC TGG GAA GAA GGT CCC AAT AGT TTT CAA CCG TCT GAT      390
Asn Gly Phe Leu Trp Glu Glu Gly Pro Asn Ser Phe Gln Pro Ser Asp
105                 110                 115                 120

CCT ATG CTC ACT ATG GTG GTA GAT AGT GGT TTG AAG GAT GAT TTG GTG      438
Pro Met Leu Thr Met Val Val Asp Ser Gly Leu Lys Asp Asp Leu Val
                125                 130                 135

TTG GGA GAT CCT ACT GCG CCA AGG TTT GTG TTG TGG AAT GGG AAA TTG      486
Leu Gly Asp Pro Thr Ala Pro Arg Phe Val Leu Trp Asn Gly Lys Leu
            140                 145                 150

AGG CCG GTT CCA TCG AAG CTA ACA GAC TTA CCG TTC TTT GAT TTG ATG      534
Arg Pro Val Pro Ser Lys Leu Thr Asp Leu Pro Phe Phe Asp Leu Met
        155                 160                 165

AGT ATT GGT GGG AAG ATT AGA GCT GGT TTT GGT GCA CTT GGC ATT CGA      582
Ser Ile Gly Gly Lys Ile Arg Ala Gly Phe Gly Ala Leu Gly Ile Arg
    170                 175                 180

CCG TCA CCT CCA GGT CGT GAA GAA TCT GTG GAG GAG TTT GTA CGG CGT      630
Pro Ser Pro Pro Gly Arg Glu Glu Ser Val Glu Glu Phe Val Arg Arg
185                 190                 195                 200

AAC CTC GGT GAT GAG GTT TTT GAG CGC CTG ATT GAA CCG TTT TGT TCA      678
Asn Leu Gly Asp Glu Val Phe Glu Arg Leu Ile Glu Pro Phe Cys Ser
```

```
                    205                 210                 215

GGT GTT TAT GCT GGT GAT CCT TCA AAA CTG AGC ATG AAA GCA GCG TTT        726
Gly Val Tyr Ala Gly Asp Pro Ser Lys Leu Ser Met Lys Ala Ala Phe
            220                 225                 230

GGG AAG GTT TGG AAA CTA GAG CAA AAT GGT GGA AGC ATA ATA GGT GGT        774
Gly Lys Val Trp Lys Leu Glu Gln Asn Gly Gly Ser Ile Ile Gly Gly
        235                 240                 245

ACT TTT AAG GCA ATT CAG GAG AGG AAA AAC GCT CCC AAG GCA GAA CGA        822
Thr Phe Lys Ala Ile Gln Glu Arg Lys Asn Ala Pro Lys Ala Glu Arg
    250                 255                 260

GAC CCG CGC CTG CCA AAA CCA CAG GGC CAA ACA GTT GGT TCT TTC AGG        870
Asp Pro Arg Leu Pro Lys Pro Gln Gly Gln Thr Val Gly Ser Phe Arg
265                 270                 275                 280

AAG GGA CTT CGA ATG TTG CCA GAA GCA ATA TCT GCA AGA TTA GGT AGC        918
Lys Gly Leu Arg Met Leu Pro Glu Ala Ile Ser Ala Arg Leu Gly Ser
                285                 290                 295

AAA GTT AAG TTG TCT TGG AAG CTC TCA GGT ATC ACT AAG CTG GAG AGC        966
Lys Val Lys Leu Ser Trp Lys Leu Ser Gly Ile Thr Lys Leu Glu Ser
            300                 305                 310

GGA GGA TAC AAC TTA ACA TAT GAG ACT CCA GAT GGT TTA GTT TCC GTG       1014
Gly Gly Tyr Asn Leu Thr Tyr Glu Thr Pro Asp Gly Leu Val Ser Val
        315                 320                 325

CAG AGC AAA AGT GTT GTA ATG ACG GTG CCA TCT CAT GTT GCA AGT GGT       1062
Gln Ser Lys Ser Val Val Met Thr Val Pro Ser His Val Ala Ser Gly
    330                 335                 340

CTC TTG CGC CCT CTT TCT GAA TCT GCT GCA AAT GCA CTC TCA AAA CTA       1110
Leu Leu Arg Pro Leu Ser Glu Ser Ala Ala Asn Ala Leu Ser Lys Leu
345                 350                 355                 360

TAT TAC CCA CCA GTT GCA GCA GTA TCT ATC TCG TAC CCG AAA GAA GCA       1158
Tyr Tyr Pro Pro Val Ala Ala Val Ser Ile Ser Tyr Pro Lys Glu Ala
                365                 370                 375

ATC CGA ACA GAA TGT TTG ATA GAT GGT GAA CTA AAG GGT TTT GGG CAA       1206
Ile Arg Thr Glu Cys Leu Ile Asp Gly Glu Leu Lys Gly Phe Gly Gln
            380                 385                 390

TTG CAT CCA CGC ACG CAA GGA GTT GAA ACA TTA GGA ACT ATC TAC AGC       1254
Leu His Pro Arg Thr Gln Gly Val Glu Thr Leu Gly Thr Ile Tyr Ser
        395                 400                 405

TCC TCA CTC TTT CCA AAT CGC GCA CCG CCC GGA AGA ATT TTG CTG TTG       1302
Ser Ser Leu Phe Pro Asn Arg Ala Pro Pro Gly Arg Ile Leu Leu Leu
    410                 415                 420

AAC TAC ATT GGC GGG TCT ACA AAC ACC GGA ATT CTG TCC AAG TCT GAA       1350
Asn Tyr Ile Gly Gly Ser Thr Asn Thr Gly Ile Leu Ser Lys Ser Glu
425                 430                 435                 440

GGT GAG TTA GTG GAA GCA GTT GAC AGA GAT TTG AGG AAA ATG CTA ATT       1398
Gly Glu Leu Val Glu Ala Val Asp Arg Asp Leu Arg Lys Met Leu Ile
                445                 450                 455

AAG CCT AAT TCG ACC GAT CCA CTT AAA TTA GGA GTT AGG GTA TGG CCT       1446
```

```
Lys Pro Asn Ser Thr Asp Pro Leu Lys Leu Gly Val Arg Val Trp Pro
            460                 465                 470

CAA GCC ATT CCT CAG TTT CTA GTT GGT CAC TTT GAT ATC CTT GAC ACG      1494
Gln Ala Ile Pro Gln Phe Leu Val Gly His Phe Asp Ile Leu Asp Thr
        475                 480                 485

GCT AAA TCA TCT CTA ACG TCT TCG GGC TAC GAA GGG CTA TTT TTG GGT      1542
Ala Lys Ser Ser Leu Thr Ser Ser Gly Tyr Glu Gly Leu Phe Leu Gly
    490                 495                 500

GGC AAT TAC GTC GCT GGT GTA GCC TTA GGC CGG TGT GTA GAA GGC GCA      1590
Gly Asn Tyr Val Ala Gly Val Ala Leu Gly Arg Cys Val Glu Gly Ala
505                 510                 515                 520

TAT GAA ACC GCG ATT GAG GTC AAC AAC TTC ATG TCA CGG TAC GCT TAC      1638
Tyr Glu Thr Ala Ile Glu Val Asn Asn Phe Met Ser Arg Tyr Ala Tyr
                525                 530                 535

AAG TAAATGTAAA ACATTAAATC TCCCAGCTTG CGTGAGTTTT ATTAAATTTT           1691
Lys

TTGAGATATC CAAAAAAAAA AAAAAAAA                                       1719
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 537 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Glu Leu Ser Leu Leu Arg Pro Thr Thr Gln Ser Leu Leu Pro Ser
 1               5                  10                  15

Phe Ser Lys Pro Asn Leu Arg Leu Asn Val Tyr Lys Pro Leu Arg Leu
            20                  25                  30

Arg Cys Ser Val Ala Gly Gly Pro Thr Val Gly Ser Ser Lys Ile Glu
        35                  40                  45

Gly Gly Gly Gly Thr Thr Ile Thr Thr Asp Cys Val Ile Val Gly Gly
    50                  55                  60

Gly Ile Ser Gly Leu Cys Ile Ala Gln Ala Leu Ala Thr Lys His Pro
65                  70                  75                  80

Asp Ala Ala Pro Asn Leu Ile Val Thr Glu Ala Lys Asp Arg Val Gly
                85                  90                  95

Gly Asn Ile Ile Thr Arg Glu Glu Asn Gly Phe Leu Trp Glu Glu Gly
            100                 105                 110

Pro Asn Ser Phe Gln Pro Ser Asp Pro Met Leu Thr Met Val Val Asp
        115                 120                 125

Ser Gly Leu Lys Asp Asp Leu Val Leu Gly Asp Pro Thr Ala Pro Arg
    130                 135                 140
```

```
Phe Val Leu Trp Asn Gly Lys Leu Arg Pro Val Pro Ser Lys Leu Thr
145                 150                 155                 160

Asp Leu Pro Phe Phe Asp Leu Met Ser Ile Gly Gly Lys Ile Arg Ala
                165                 170                 175

Gly Phe Gly Ala Leu Gly Ile Arg Pro Ser Pro Pro Gly Arg Glu Glu
            180                 185                 190

Ser Val Glu Glu Phe Val Arg Arg Asn Leu Gly Asp Glu Val Phe Glu
        195                 200                 205

Arg Leu Ile Glu Pro Phe Cys Ser Gly Val Tyr Ala Gly Asp Pro Ser
    210                 215                 220

Lys Leu Ser Met Lys Ala Ala Phe Gly Lys Val Trp Lys Leu Glu Gln
225                 230                 235                 240

Asn Gly Gly Ser Ile Ile Gly Gly Thr Phe Lys Ala Ile Gln Glu Arg
                245                 250                 255

Lys Asn Ala Pro Lys Ala Glu Arg Asp Pro Arg Leu Pro Lys Pro Gln
            260                 265                 270

Gly Gln Thr Val Gly Ser Phe Arg Lys Gly Leu Arg Met Leu Pro Glu
        275                 280                 285

Ala Ile Ser Ala Arg Leu Gly Ser Lys Val Lys Leu Ser Trp Lys Leu
    290                 295                 300

Ser Gly Ile Thr Lys Leu Glu Ser Gly Gly Tyr Asn Leu Thr Tyr Gln
305                 310                 315                 320

Thr Pro Asp Gly Leu Val Ser Val Gln Ser Lys Ser Val Val Met Thr
                325                 330                 335

Val Pro Ser His Val Ala Ser Gly Leu Leu Arg Pro Leu Ser Glu Ser
            340                 345                 350

Ala Ala Asn Ala Leu Ser Lys Leu Tyr Tyr Pro Pro Val Ala Ala Val
        355                 360                 365

Ser Ile Ser Tyr Pro Lys Glu Ala Ile Arg Thr Glu Cys Leu Ile Asp
    370                 375                 380

Gly Glu Leu Lys Gly Phe Gly Gln Leu His Pro Arg Thr Gln Gly Val
385                 390                 395                 400

Glu Thr Leu Gly Thr Ile Tyr Ser Ser Ser Leu Phe Pro Asn Arg Ala
                405                 410                 415

Pro Pro Gly Arg Ile Leu Leu Leu Asn Tyr Ile Gly Gly Ser Thr Asn
            420                 425                 430

Thr Gly Ile Leu Ser Lys Ser Glu Gly Glu Leu Val Glu Ala Val Asp
        435                 440                 445

Arg Asp Leu Arg Lys Met Leu Ile Lys Pro Asn Ser Thr Asp Pro Leu
    450                 455                 460

Lys Leu Gly Val Arg Val Trp Pro Gln Ala Ile Pro Gln Phe Leu Val
465                 470                 475                 480
```

```
Gly His Phe Asp Ile Leu Asp Thr Ala Lys Ser Ser Leu Thr Ser Ser
                485                 490                 495

Gly Tyr Glu Gly Leu Phe Leu Gly Gly Asn Tyr Val Ala Gly Val Ala
            500                 505                 510

Leu Gly Arg Cys Val Glu Gly Ala Tyr Glu Thr Ala Ile Glu Val Asn
        515                 520                 525

Asn Phe Met Ser Arg Tyr Ala Tyr Lys
    530                 535
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1738 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 70..1596
(D) OTHER INFORMATION: /note= "Arabidopsis protox-2 cDNA; sequence from pWDC-1"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
TTTTTTACTT ATTTCCGTCA CTGCTTTCGA CTGGTCAGAG ATTTTGACTC TGAATTGTTG      60

CAGATAGCA ATG GCG TCT GGA GCA GTA GCA GAT CAT CAA ATT GAA GCG         108
          Met Ala Ser Gly Ala Val Ala Asp His Gln Ile Glu Ala
            1               5                  10

GTT TCA GGA AAA AGA GTC GCA GTC GTA GGT GCA GGT GTA AGT GGA CTT       156
Val Ser Gly Lys Arg Val Ala Val Val Gly Ala Gly Val Ser Gly Leu
     15                  20                  25

GCG GCG GCT TAC AAG TTG AAA TCG AGG GGT TTG AAT GTG ACT GTG TTT       204
Ala Ala Ala Tyr Lys Leu Lys Ser Arg Gly Leu Asn Val Thr Val Phe
 30                  35                  40                  45

GAA GCT GAT GGA AGA GTA GGT GGG AAG TTG AGA AGT GTT ATG CAA AAT       252
Glu Ala Asp Gly Arg Val Gly Gly Lys Leu Arg Ser Val Met Gln Asn
                 50                  55                  60

GGT TTG ATT TGG GAT GAA GGA GCA AAC ACC ATG ACT GAG GCT GAG CCA       300
Gly Leu Ile Trp Asp Glu Gly Ala Asn Thr Met Thr Glu Ala Glu Pro
             65                  70                  75

GAA GTT GGG AGT TTA CTT GAT GAT CTT GGG CTT CGT GAG AAA CAA CAA       348
Glu Val Gly Ser Leu Leu Asp Asp Leu Gly Leu Arg Glu Lys Gln Gln
         80                  85                  90
```

```
TTT CCA ATT TCA CAG AAA AAG CGG TAT ATT GTG CGG AAT GGT GTA CCT      396
Phe Pro Ile Ser Gln Lys Lys Arg Tyr Ile Val Arg Asn Gly Val Pro
     95                 100                 105

GTG ATG CTA CCT ACC AAT CCC ATA GAG CTG GTC ACA AGT AGT GTG CTC      444
Val Met Leu Pro Thr Asn Pro Ile Glu Leu Val Thr Ser Ser Val Leu
110                 115                 120                 125

TCT ACC CAA TCT AAG TTT CAA ATC TTG TTG GAA CCA TTT TTA TGG AAG      492
Ser Thr Gln Ser Lys Phe Gln Ile Leu Leu Glu Pro Phe Leu Trp Lys
                130                 135                 140

AAA AAG TCC TCA AAA GTC TCA GAT GCA TCT GCT GAA GAA AGT GTA AGC      540
Lys Lys Ser Ser Lys Val Ser Asp Ala Ser Ala Glu Glu Ser Val Ser
            145                 150                 155

GAG TTC TTT CAA CGC CAT TTT GGA CAA GAG GTT GTT GAC TAT CTC ATC      588
Glu Phe Phe Gln Arg His Phe Gly Gln Glu Val Val Asp Tyr Leu Ile
        160                 165                 170

GAC CCT TTT GTT GGT GGA ACA AGT GCT GCG GAC CCT GAT TCC CTT TCA      636
Asp Pro Phe Val Gly Gly Thr Ser Ala Ala Asp Pro Asp Ser Leu Ser
    175                 180                 185

ATG AAG CAT TCT TTC CCA GAT CTC TGG AAT GTA GAG AAA AGT TTT GGC      684
Met Lys His Ser Phe Pro Asp Leu Trp Asn Val Glu Lys Ser Phe Gly
190                 195                 200                 205

TCT ATT ATA GTC GGT GCA ATC AGA ACA AAG TTT GCT GCT AAA GGT GGT      732
Ser Ile Ile Val Gly Ala Ile Arg Thr Lys Phe Ala Ala Lys Gly Gly
                210                 215                 220

AAA AGT AGA GAC ACA AAG AGT TCT CCT GGC ACA AAA AAG GGT TCG CGT      780
Lys Ser Arg Asp Thr Lys Ser Ser Pro Gly Thr Lys Lys Gly Ser Arg
            225                 230                 235

GGG TCA TTC TCT TTT AAG GGG GGA ATG CAG ATT CTT CCT GAT ACG TTG      828
Gly Ser Phe Ser Phe Lys Gly Gly Met Gln Ile Leu Pro Asp Thr Leu
        240                 245                 250

TGC AAA AGT CTC TCA CAT GAT GAG ATC AAT TTA GAC TCC AAG GTA CTC      876
Cys Lys Ser Leu Ser His Asp Glu Ile Asn Leu Asp Ser Lys Val Leu
    255                 260                 265

TCT TTG TCT TAC AAT TCT GGA TCA AGA CAG GAG AAC TGG TCA TTA TCT      924
Ser Leu Ser Tyr Asn Ser Gly Ser Arg Gln Glu Asn Trp Ser Leu Ser
270                 275                 280                 285

TGT GTT TCG CAT AAT GAA ACG CAG AGA CAA AAC CCC CAT TAT GAT GCT      972
Cys Val Ser His Asn Glu Thr Gln Arg Gln Asn Pro His Tyr Asp Ala
                290                 295                 300

GTA ATT ATG ACG GCT CCT CTG TGC AAT GTG AAG GAG ATG AAG GTT ATG     1020
Val Ile Met Thr Ala Pro Leu Cys Asn Val Lys Glu Met Lys Val Met
            305                 310                 315

AAA GGA GGA CAA CCC TTT CAG CTA AAC TTT CTC CCC GAG ATT AAT TAC     1068
Lys Gly Gly Gln Pro Phe Gln Leu Asn Phe Leu Pro Glu Ile Asn Tyr
        320                 325                 330

ATG CCC CTC TCG GTT TTA ATC ACC ACA TTC ACA AAG GAG AAA GTA AAG     1116
Met Pro Leu Ser Val Leu Ile Thr Thr Phe Thr Lys Glu Lys Val Lys
```

```
    335                         340                         345

AGA CCT CTT GAA GGC TTT GGG GTA CTC ATT CCA TCT AAG GAG CAA AAG        1164
Arg Pro Leu Glu Gly Phe Gly Val Leu Il  Pro Ser Lys Glu Gln Lys
350                     355                     360                 365

CAT GGT TTC AAA ACT CTA GGT ACA CTT TTT TCA TCA ATG ATG TTT CCA        1212
His Gly Phe Lys Thr Leu Gly Thr Leu Phe Ser Ser Met Met Phe Pro
                370                     375                     380

GAT CGT TCC CCT AGT GAC GTT CAT CTA TAT ACA ACT TTT ATT GGT GGG        1260
Asp Arg Ser Pro Ser Asp Val His Leu Tyr Thr Thr Phe Ile Gly Gly
            385                     390                     395

AGT AGG AAC CAG GAA CTA GCC AAA GCT TCC ACT GAC GAA TTA AAA CAA        1308
Ser Arg Asn Gln Glu Leu Ala Lys Ala Ser Thr Asp Glu Leu Lys Gln
        400                     405                     410

GTT GTG ACT TCT GAC CTT CAG CGA CTG TTG GGG GTT GAA GGT GAA CCC        1356
Val Val Thr Ser Asp Leu Gln Arg Leu Leu Gly Val Glu Gly Glu Pro
    415                         420                     425

GTG TCT GTC AAC CAT TAC TAT TGG AGG AAA GCA TTC CCG TTG TAT GAC        1404
Val Ser Val Asn His Tyr Tyr Trp Arg Lys Ala Phe Pro Leu Tyr Asp
430                     435                     440                 445

AGC AGC TAT GAC TCA GTC ATG GAA GCA ATT GAC AAG ATG GAG AAT GAT        1452
Ser Ser Tyr Asp Ser Val Met Glu Ala Ile Asp Lys Met Glu Asn Asp
                450                     455                     460

CTA CCT GGG TTC TTC TAT GCA GGT AAT CAT CGA GGG GGG CTC TCT GTT        1500
Leu Pro Gly Phe Phe Tyr Ala Gly Asn His Arg Gly Gly Leu Ser Val
            465                     470                     475

GGG AAA TCA ATA GCA TCA GGT TGC AAA GCA GCT GAC CTT GTG ATC TCA        1548
Gly Lys Ser Ile Ala Ser Gly Cys Lys Ala Ala Asp Leu Val Ile Ser
        480                     485                     490

TAC CTG GAG TCT TGC TCA AAT GAC AAG AAA CCA AAT GAC AGC TTA TAACATTGTC    1603
Tyr Leu Glu Ser Cys Ser Asn Asp Lys Lys Pro Asn Asp Ser Leu
    495                         500                     505

AAGGTTCGTC CCTTTTTATC ACTTACTTTG TAAACTTGTA AAATGCAACA AGCCGCCGTG      1663

CGATTAGCCA ACAACTCAGC AAAACCCAGA TTCTCATAAG GCTCACTAAT TCCAGAATAA      1723

ACTATTTATG TAAAA                                                       1738
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 508 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Ala Ser Gly Ala Val Ala Asp His Gln Ile Glu Ala Val Ser Gly
1               5                   10                  15
```

```
Lys Arg Val Ala Val Val Gly Ala Gly Val Ser Gly Leu Ala Ala Ala
            20                  25                  30

Tyr Lys Leu Lys Ser Arg Gly Leu Asn Val Thr Val Phe Glu Ala Asp
        35                  40                  45

Gly Arg Val Gly Gly Lys Leu Arg Ser Val Met Gln Asn Gly Leu Ile
    50                  55                  60

Trp Asp Glu Gly Ala Asn Thr Met Thr Glu Ala Glu Pro Glu Val Gly
65                  70                  75                  80

Ser Leu Leu Asp Asp Leu Gly Leu Arg Glu Lys Gln Gln Phe Pro Ile
                85                  90                  95

Ser Gln Lys Lys Arg Tyr Ile Val Arg Asn Gly Val Pro Val Met Leu
            100                 105                 110

Pro Thr Asn Pro Ile Glu Leu Val Thr Ser Ser Val Leu Ser Thr Gln
        115                 120                 125

Ser Lys Phe Gln Ile Leu Leu Glu Pro Phe Leu Trp Lys Lys Lys Ser
    130                 135                 140

Ser Lys Val Ser Asp Ala Ser Ala Glu Glu Ser Val Ser Glu Phe Phe
145                 150                 155                 160

Gln Arg His Phe Gly Gln Glu Val Val Asp Tyr Leu Ile Asp Pro Phe
                165                 170                 175

Val Gly Gly Thr Ser Ala Ala Asp Pro Asp Ser Leu Ser Met Lys His
            180                 185                 190

Ser Phe Pro Asp Leu Trp Asn Val Glu Lys Ser Phe Gly Ser Ile Ile
        195                 200                 205

Val Gly Ala Ile Arg Thr Lys Phe Ala Ala Lys Gly Gly Lys Ser Arg
    210                 215                 220

Asp Thr Lys Ser Ser Pro Gly Thr Lys Lys Gly Ser Arg Gly Ser Phe
225                 230                 235                 240

Ser Phe Lys Gly Gly Met Gln Ile Leu Pro Asp Thr Leu Cys Lys Ser
                245                 250                 255

Leu Ser His Asp Glu Ile Asn Leu Asp Ser Lys Val Leu Ser Leu Ser
            260                 265                 270

Tyr Asn Ser Gly Ser Arg Gln Glu Asn Trp Ser Leu Ser Cys Val Ser
        275                 280                 285

His Asn Glu Thr Gln Arg Gln Asn Pro His Tyr Asp Ala Val Ile Met
    290                 295                 300

Thr Ala Pro Leu Cys Asn Val Lys Glu Met Lys Val Met Lys Gly Gly
305                 310                 315                 320

Gln Pro Phe Gln Leu Asn Phe Leu Pro Glu Ile Asn Tyr Met Pro Leu
                325                 330                 335

Ser Val Leu Ile Thr Thr Phe Thr Lys Glu Lys Val Lys Arg Pro Leu
            340                 345                 350
```

```
Glu Gly Phe Gly Val Leu Ile Pro Ser Lys Glu Gln Lys His Gly Phe
        355                 360                 365

Lys Thr Leu Gly Thr Leu Phe Ser Ser Met Met Phe Pro Asp Arg Ser
    370                 375                 380

Pro Ser Asp Val His Leu Tyr Thr Thr Phe Ile Gly Gly Ser Arg Asn
385                 390                 395                 400

Gln Glu Leu Ala Lys Ala Ser Thr Asp Glu Leu Lys Gln Val Val Thr
                405                 410                 415

Ser Asp Leu Gln Arg Leu Leu Gly Val Glu Gly Glu Pro Val Ser Val
            420                 425                 430

Asn His Tyr Tyr Trp Arg Lys Ala Phe Pro Leu Tyr Asp Ser Ser Tyr
        435                 440                 445

Asp Ser Val Met Glu Ala Ile Asp Lys Met Glu Asn Asp Leu Pro Gly
    450                 455                 460

Phe Phe Tyr Ala Gly Asn His Arg Gly Gly Leu Ser Val Gly Lys Ser
465                 470                 475                 480

Ile Ala Ser Gly Cys Lys Ala Ala Asp Leu Val Ile Ser Tyr Leu Glu
                485                 490                 495

Ser Cys Ser Asn Asp Lys Lys Pro Asn Asp Ser Leu
            500                 505
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1698 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
- (A) NAME/KEY: CDS
- (B) LOCATION: 2..1453
- (D) OTHER INFORMATION: /note= "Maize protox-1 cDNA (not full-length); sequence from pWDC-4"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
G AAT TCG GCG GAC TGC GTC GTG GTG GGC GGA GGC ATC AGT GGC CTC      46
  Asn Ser Ala Asp Cys Val Val Val Gly Gly Gly Ile Ser Gly Leu
  1               5                   10                  15

TGC ACC GCG CAG GCG CTG GCC ACG CGG CAC GGC GTC GGG GAC GTG CTT    94
Cys Thr Ala Gln Ala Leu Ala Thr Arg His Gly Val Gly Asp Val Leu
                20                  25                  30
```

```
GTC ACG GAG GCC CGC GCC CGC CCC GGC GGC AAC ATT ACC ACC GTC GAG        142
Val Thr Glu Ala Arg Ala Arg Pro Gly Gly Asn Il  Thr Thr Val Glu
            35                  40                  45

CGC CCC GAG GAA GGG TAC CTC TGG GAG GAG GGT CCC AAC AGC TTC CAG        190
Arg Pro Glu Glu Gly Tyr Leu Trp Glu Glu Gly Pro Asn Ser Phe Gln
            50                  55                  60

CCC TCC GAC CCC GTT CTC ACC ATG GCC GTG GAC AGC GGA CTG AAG GAT        238
Pro Ser Asp Pro Val Leu Thr Met Ala Val Asp Ser Gly Leu Lys Asp
    65                  70                  75

GAC TTG GTT TTT GGG GAC CCA AAC GCG CCG CGT TTC GTG CTG TGG GAG        286
Asp Leu Val Phe Gly Asp Pro Asn Ala Pro Arg Phe Val Leu Trp Glu
80                  85                  90                  95

GGG AAG CTG AGG CCC GTG CCA TCC AAG CCC GCC GAC CTC CCG TTC TTC        334
Gly Lys Leu Arg Pro Val Pro Ser Lys Pro Ala Asp Leu Pro Phe Phe
                100                 105                 110

GAT CTC ATG AGC ATC CCA GGG AAG CTC AGG GCC GGT CTA GGC GCG CTT        382
Asp Leu Met Ser Ile Pro Gly Lys Leu Arg Ala Gly Leu Gly Ala Leu
            115                 120                 125

GGC ATC CGC CCG CCT CCT CCA GGC CGC GAA GAG TCA GTG GAG GAG TTC        430
Gly Ile Arg Pro Pro Pro Pro Gly Arg Glu Glu Ser Val Glu Glu Phe
        130                 135                 140

GTG CGC CGC AAC CTC GGT GCT GAG GTC TTT GAG CGC CTC ATT GAG CCT        478
Val Arg Arg Asn Leu Gly Ala Glu Val Phe Glu Arg Leu Ile Glu Pro
    145                 150                 155

TTC TGC TCA GGT GTC TAT GCT GGT GAT CCT TCT AAG CTC AGC ATG AAG        526
Phe Cys Ser Gly Val Tyr Ala Gly Asp Pro Ser Lys Leu Ser Met Lys
160                 165                 170                 175

GCT GCA TTT GGG AAG GTT TGG CGG TTG GAA GAA ACT GGA GGT AGT ATT        574
Ala Ala Phe Gly Lys Val Trp Arg Leu Glu Glu Thr Gly Gly Ser Ile
                180                 185                 190

ATT GGT GGA ACC ATC AAG ACA ATT CAG GAG AGG AGC AAG AAT CCA AAA        622
Ile Gly Gly Thr Ile Lys Thr Ile Gln Glu Arg Ser Lys Asn Pro Lys
            195                 200                 205

CCA CCG AGG GAT GCC CGC CTT CCG AAG CCA AAA GGG CAG ACA GTT GCA        670
Pro Pro Arg Asp Ala Arg Leu Pro Lys Pro Lys Gly Gln Thr Val Ala
        210                 215                 220

TCT TTC AGG AAG GGT CTT GCC ATG CTT CCA AAT GCC ATT ACA TCC AGC        718
Ser Phe Arg Lys Gly Leu Ala Met Leu Pro Asn Ala Ile Thr Ser Ser
    225                 230                 235

TTG GGT AGT AAA GTC AAA CTA TCA TGG AAA CTC ACG AGC ATT ACA AAA        766
Leu Gly Ser Lys Val Lys Leu Ser Trp Lys Leu Thr Ser Ile Thr Lys
240                 245                 250                 255

TCA GAT GAC AAG GGA TAT GTT TTG GAG TAT GAA ACG CCA GAA GGG GTT        814
Ser Asp Asp Lys Gly Tyr Val Leu Glu Tyr Glu Thr Pro Glu Gly Val
                260                 265                 270

GTT TCG GTG CAG GCT AAA AGT GTT ATC ATG ACT ATT CCA TCA TAT GTT        862
Val Ser Val Gln Ala Lys Ser Val Ile Met Thr Ile Pro Ser Tyr Val
            275                 280                 285
```

```
GCT AGC AAC ATT TTG CGT CCA CTT TCA AGC GAT GCT GCA GAT GCT CTA        910
Ala Ser Asn Ile Leu Arg Pro Leu Ser Ser Asp Ala Ala Asp Ala Leu
        290                 295                 300

TCA AGA TTC TAT TAT CCA CCG GTT GCT GCT GTA ACT GTT TCG TAT CCA        958
Ser Arg Phe Tyr Tyr Pro Pro Val Ala Ala Val Thr Val Ser Tyr Pro
    305                 310                 315

AAG GAA GCA ATT AGA AAA GAA TGC TTA ATT GAT GGG GAA CTC CAG GGC       1006
Lys Glu Ala Ile Arg Lys Glu Cys Leu Ile Asp Gly Glu Leu Gln Gly
320                 325                 330                 335

TTT GGC CAG TTG CAT CCA CGT AGT CAA GGA GTT GAG ACA TTA GGA ACA       1054
Phe Gly Gln Leu His Pro Arg Ser Gln Gly Val Glu Thr Leu Gly Thr
                340                 345                 350

ATA TAC AGT TCC TCA CTC TTT CCA AAT CGT GCT CCT GAC GGT AGG GTG       1102
Ile Tyr Ser Ser Ser Leu Phe Pro Asn Arg Ala Pro Asp Gly Arg Val
            355                 360                 365

TTA CTT CTA AAC TAC ATA GGA GGT GCT ACA AAC ACA GGA ATT GTT TCC       1150
Leu Leu Leu Asn Tyr Ile Gly Gly Ala Thr Asn Thr Gly Ile Val Ser
        370                 375                 380

AAG ACT GAA AGT GAG CTG GTC GAA GCA GTT GAC CGT GAC CTC CGA AAA       1198
Lys Thr Glu Ser Glu Leu Val Glu Ala Val Asp Arg Asp Leu Arg Lys
    385                 390                 395

ATG CTT ATA AAT TCT ACA GCA GTG GAC CCT TTA GTC CTT GGT GTT CGA       1246
Met Leu Ile Asn Ser Thr Ala Val Asp Pro Leu Val Leu Gly Val Arg
400                 405                 410                 415

GTT TGG CCA CAA GCC ATA CCT CAG TTC CTG GTA GGA CAT CTT GAT CTT       1294
Val Trp Pro Gln Ala Ile Pro Gln Phe Leu Val Gly His Leu Asp Leu
                420                 425                 430

CTG GAA GCC GCA AAA GCT GCC CTG GAC CGA GGT GGC TAC GAT GGG CTG       1342
Leu Glu Ala Ala Lys Ala Ala Leu Asp Arg Gly Gly Tyr Asp Gly Leu
            435                 440                 445

TTC CTA GGA GGG AAC TAT GTT GCA GGA GTT GCC CTG GGC AGA TGC GTT       1390
Phe Leu Gly Gly Asn Tyr Val Ala Gly Val Ala Leu Gly Arg Cys Val
        450                 455                 460

GAG GGC GCG TAT GAA AGT GCC TCG CAA ATA TCT GAC TTC TTG ACC AAG       1438
Glu Gly Ala Tyr Glu Ser Ala Ser Gln Ile Ser Asp Phe Leu Thr Lys
    465                 470                 475

TAT GCC TAC AAG TGATGAAAGA AGTGGAGCGC TACTTGTTAA TCGTTTATGT           1490
Tyr Ala Tyr Lys
480

TGCATAGATG AGGTGCCTCC GGGGAAAAAA AAGCTTGAAT AGTATTTTTT ATTCTTATTT     1550

TGTAAATTGC ATTTCTGTTC TTTTTTCTAT CAGTAATTAG TTATATTTTA GTTCTGTAGG     1610

AGATTGTTCT GTTCACTGCC CTTCAAAAGA AATTTTATTT TTCATTCTTT TATGAGAGCT     1670

GTGCTACTTA AAAAAAAAAA AAAAAAAA                                        1698
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 483 amino acids
   (B) TYPE: amino acid
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Asn Ser Ala Asp Cys Val Val Val Gly Gly Gly Ile Ser Gly Leu Cys
1               5                   10                  15

Thr Ala Gln Ala Leu Ala Thr Arg His Gly Val Gly Asp Val Leu Val
            20                  25                  30

Thr Glu Ala Arg Ala Arg Pro Gly Gly Asn Ile Thr Thr Val Glu Arg
        35                  40                  45

Pro Glu Glu Gly Tyr Leu Trp Gln Glu Gly Pro Asn Ser Phe Gln Pro
    50                  55                  60

Ser Asp Pro Val Leu Thr Met Ala Val Asp Ser Gly Leu Lys Asp Asp
65                  70                  75                  80

Leu Val Phe Gly Asp Pro Asn Ala Pro Arg Phe Val Leu Trp Glu Gly
                85                  90                  95

Lys Leu Arg Pro Val Pro Ser Lys Pro Ala Asp Leu Pro Phe Phe Asp
            100                 105                 110

Leu Met Ser Ile Pro Gly Lys Leu Arg Ala Gly Leu Gly Ala Leu Gly
        115                 120                 125

Ile Arg Pro Pro Pro Pro Gly Arg Glu Glu Ser Val Glu Glu Phe Val
    130                 135                 140

Arg Arg Asn Leu Gly Ala Glu Val Phe Glu Arg Leu Ile Glu Pro Phe
145                 150                 155                 160

Cys Ser Gly Val Tyr Ala Gly Asp Pro Ser Lys Leu Ser Met Lys Ala
                165                 170                 175

Ala Phe Gly Lys Val Trp Arg Leu Glu Glu Thr Gly Gly Ser Ile Ile
            180                 185                 190

Gly Gly Thr Ile Lys Thr Ile Gln Glu Arg Ser Lys Asn Pro Lys Pro
        195                 200                 205

Pro Arg Asp Ala Arg Leu Pro Lys Pro Lys Gly Gln Thr Val Ala Ser
    210                 215                 220

Phe Arg Lys Gly Leu Ala Met Leu Pro Asn Ala Ile Thr Ser Ser Leu
225                 230                 235                 240

Gly Ser Lys Val Lys Leu Ser Trp Lys Leu Thr Ser Ile Thr Lys Ser
                245                 250                 255

Asp Asp Lys Gly Tyr Val Leu Glu Tyr Glu Thr Pro Glu Gly Val Val
            260                 265                 270

Ser Val Gln Ala Lys Ser Val Ile Met Thr Ile Pro Ser Tyr Val Ala
        275                 280                 285
```

```
Ser Asn Ile Leu Arg Pro Leu Ser Ser Asp Ala Ala Asp Ala Leu Ser
    290                 295                 300

Arg Phe Tyr Tyr Pro Pro Val Ala Ala Val Thr Val Ser Tyr Pro Lys
305                 310                 315                 320

Glu Ala Ile Arg Lys Glu Cys Leu Ile Asp Gly Glu Leu Gln Gly Phe
                325                 330                 335

Gly Gln Leu His Pro Arg Ser Gln Gly Val Glu Thr Leu Gly Thr Ile
            340                 345                 350

Tyr Ser Ser Ser Leu Phe Pro Asn Arg Ala Pro Asp Gly Arg Val Leu
        355                 360                 365

Leu Leu Asn Tyr Ile Gly Gly Ala Thr Asn Thr Gly Ile Val Ser Lys
    370                 375                 380

Thr Glu Ser Glu Leu Val Glu Ala Val Asp Arg Asp Leu Arg Lys Met
385                 390                 395                 400

Leu Ile Asn Ser Thr Ala Val Asp Pro Leu Val Leu Gly Val Arg Val
                405                 410                 415

Trp Pro Gln Ala Ile Pro Gln Phe Leu Val Gly His Leu Asp Leu Leu
            420                 425                 430

Glu Ala Ala Lys Ala Ala Leu Asp Arg Gly Gly Tyr Asp Gly Leu Phe
        435                 440                 445

Leu Gly Gly Asn Tyr Val Ala Gly Val Ala Leu Gly Arg Cys Val Glu
    450                 455                 460

Gly Ala Tyr Glu Ser Ala Ser Gln Ile Ser Asp Phe Leu Thr Lys Tyr
465                 470                 475                 480

Ala Tyr Lys
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2061 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 64..1698
(D) OTHER INFORMATION: /note= "Maize protox-2 cDNA; sequence from pWDC-3"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
CTCTCCTACC TCCACCTCCA CCACAACAAG CAAATCCCCA TCCAGTTCCA AACCCTAACT        60

CAA ATG CTC GCT TTG ACT GCC TCA GCC TCA TCC GCT TCG TCC CAT CCT         108
    Met Leu Ala Leu Thr Ala Ser Ala Ser Ser Ala Ser Ser His Pro
     1               5                   10                  15

TAT CGC CAC GCC TCC GCG CAC ACT CGT CGC CCC CGC CTA CGT GCG GTC         156
Tyr Arg His Ala Ser Ala His Thr Arg Arg Pro Arg Leu Arg Ala Val
                 20                  25                  30

CTC GCG ATG GCG GGC TCC GAC GAC CCC CGT GCA GCG CCC GCC AGA TCG         204
Leu Ala Met Ala Gly Ser Asp Asp Pro Arg Ala Ala Pro Ala Arg Ser
             35                  40                  45

GTC GCC GTC GTC GGC GCC GGG GTC AGC GGG CTC GCG GCG GCG TAC AGG         252
Val Ala Val Val Gly Ala Gly Val Ser Gly Leu Ala Ala Ala Tyr Arg
         50                  55                  60

CTC AGA CAG AGC GGC GTG AAC GTA ACG GTG TTC GAA GCG GCC GAC AGG         300
Leu Arg Gln Ser Gly Val Asn Val Thr Val Phe Glu Ala Ala Asp Arg
     65                  70                  75

GCG GGA GGA AAG ATA CGG ACC AAT TCC GAG GGC GGG TTT GTC TGG GAT         348
Ala Gly Gly Lys Ile Arg Thr Asn Ser Glu Gly Gly Phe Val Trp Asp
 80                  85                  90                  95

GAA GGA GCT AAC ACC ATG ACA GAA GGT GAA TGG GAG GCC AGT AGA CTG         396
Glu Gly Ala Asn Thr Met Thr Glu Gly Glu Trp Glu Ala Ser Arg Leu
                 100                 105                 110

ATT GAT GAT CTT GGT CTA CAA GAC AAA CAG CAG TAT CCT AAC TCC CAA         444
Ile Asp Asp Leu Gly Leu Gln Asp Lys Gln Gln Tyr Pro Asn Ser Gln
             115                 120                 125

CAC AAG CGT TAC ATT GTC AAA GAT GGA GCA CCA GCA CTG ATT CCT TCG         492
His Lys Arg Tyr Ile Val Lys Asp Gly Ala Pro Ala Leu Ile Pro Ser
         130                 135                 140

GAT CCC ATT TCG CTA ATG AAA AGC AGT GTT CTT TCG ACA AAA TCA AAG         540
Asp Pro Ile Ser Leu Met Lys Ser Ser Val Leu Ser Thr Lys Ser Lys
     145                 150                 155

ATT GCG TTA TTT TTT GAA CCA TTT CTC TAC AAG AAA GCT AAC ACA AGA         588
Ile Ala Leu Phe Phe Glu Pro Phe Leu Tyr Lys Lys Ala Asn Thr Arg
160                 165                 170                 175

AAC TCT GGA AAA GTG TCT GAG GAG CAC TTG AGT GAG AGT GTT GGG AGC         636
Asn Ser Gly Lys Val Ser Glu Glu His Leu Ser Glu Ser Val Gly Ser
                 180                 185                 190

TTC TGT GAA CGC CAC TTT GGA AGA GAA GTT GTT GAC TAT TTT GTT GAT         684
Phe Cys Glu Arg His Phe Gly Arg Glu Val Val Asp Tyr Phe Val Asp
             195                 200                 205

CCA TTT GTA GCT GGA ACA AGT GCA GGA GAT CCA GAG TCA CTA TCT ATT         732
Pro Phe Val Ala Gly Thr Ser Ala Gly Asp Pro Glu Ser Leu Ser Ile
         210                 215                 220

CGT CAT GCA TTC CCA GCA TTG TGG AAT TTG GAA AGA AAG TAT GGT TCA         780
Arg His Ala Phe Pro Ala Leu Trp Asn Leu Glu Arg Lys Tyr Gly Ser
     225                 230                 235
```

```
GTT ATT GTT GGT GCC ATC TTG TCT AAG CTA GCA GCT AAA GGT GAT CCA      828
Val Ile Val Gly Ala Ile Leu Ser Lys Leu Ala Ala Lys Gly Asp Pro
240                 245                 250                 255

GTA AAG ACA AGA CAT GAT TCA TCA GGG AAA AGA AGG AAT AGA CGA GTG      876
Val Lys Thr Arg His Asp Ser Ser Gly Lys Arg Arg Asn Arg Arg Val
                260                 265                 270

TCG TTT TCA TTT CAT GGT GGA ATG CAG TCA CTA ATA AAT GCA CTT CAC      924
Ser Phe Ser Phe His Gly Gly Met Gln Ser Leu Ile Asn Ala Leu His
            275                 280                 285

AAT GAA GTT GGA GAT GAT AAT GTG AAG CTT GGT ACA GAA GTG TTG TCA      972
Asn Glu Val Gly Asp Asp Asn Val Lys Leu Gly Thr Glu Val Leu Ser
        290                 295                 300

TTG GCA TGT ACA TTT GAT GGA GTT CCT GCA CTA GGC AGG TGG TCA ATT     1020
Leu Ala Cys Thr Phe Asp Gly Val Pro Ala Leu Gly Arg Trp Ser Ile
    305                 310                 315

TCT GTT GAT TCG AAG GAT AGC GGT GAC AAG GAC CTT GCT AGT AAC CAA     1068
Ser Val Asp Ser Lys Asp Ser Gly Asp Lys Asp Leu Ala Ser Asn Gln
320                 325                 330                 335

ACC TTT GAT GCT GTT ATA ATG ACA GCT CCA TTG TCA AAT GTC CGG AGG     1116
Thr Phe Asp Ala Val Ile Met Thr Ala Pro Leu Ser Asn Val Arg Arg
                340                 345                 350

ATG AAG TTC ACC AAA GGT GGA GCT CCG GTT GTT CTT GAC TTT CTT CCT     1164
Met Lys Phe Thr Lys Gly Gly Ala Pro Val Val Leu Asp Phe Leu Pro
            355                 360                 365

AAG ATG GAT TAT CTA CCA CTA TCT CTC ATG GTG ACT GCT TTT AAG AAG     1212
Lys Met Asp Tyr Leu Pro Leu Ser Leu Met Val Thr Ala Phe Lys Lys
        370                 375                 380

GAT GAT GTC AAG AAA CCT CTG GAA GGA TTT GGG GTC TTA ATA CCT TAC     1260
Asp Asp Val Lys Lys Pro Leu Glu Gly Phe Gly Val Leu Ile Pro Tyr
    385                 390                 395

AAG GAA CAG CAA AAA CAT GGT CTG AAA ACC CTT GGG ACT CTC TTT TCC     1308
Lys Glu Gln Gln Lys His Gly Leu Lys Thr Leu Gly Thr Leu Phe Ser
400                 405                 410                 415

TCA ATG ATG TTC CCA GAT CGA GCT CCT GAT GAC CAA TAT TTA TAT ACA     1356
Ser Met Met Phe Pro Asp Arg Ala Pro Asp Asp Gln Tyr Leu Tyr Thr
                420                 425                 430

ACA TTT GTT GGG GGT AGC CAC AAT AGA GAT CTT GCT GGA GCT CCA ACG     1404
Thr Phe Val Gly Gly Ser His Asn Arg Asp Leu Ala Gly Ala Pro Thr
            435                 440                 445

TCT ATT CTG AAA CAA CTT GTG ACC TCT GAC CTT AAA AAA CTC TTG GGC     1452
Ser Ile Leu Lys Gln Leu Val Thr Ser Asp Leu Lys Lys Leu Leu Gly
        450                 455                 460

GTA GAG GGG CAA CCA ACT TTT GTC AAG CAT GTA TAC TGG GGA AAT GCT     1500
Val Glu Gly Gln Pro Thr Phe Val Lys His Val Tyr Trp Gly Asn Ala
    465                 470                 475

TTT CCT TTG TAT GGC CAT GAT TAT AGT TCT GTA TTG GAA GCT ATA GAA     1548
Phe Pro Leu Tyr Gly His Asp Tyr Ser Ser Val Leu Glu Ala Ile Glu
```

480 485 490 495

AAG ATG GAG AAA AAC CTT CCA GGG TTC TTC TAC GCA GGA AAT AGC AAG 1596
Lys Met Glu Lys Asn Leu Pro Gly Phe Phe Tyr Ala Gly Asn Ser Lys
500 505 510

GAT GGG CTT GCT GTT GGA AGT GTT ATA GCT TCA GGA AGC AAG GCT GCT 1644
Asp Gly Leu Ala Val Gly Ser Val Ile Ala Ser Gly Ser Lys Ala Ala
515 520 525

GAC CTT GCA ATC TCA TAT CTT GAA TCT CAC ACC AAG CAT AAT AAT TCA 1692
Asp Leu Ala Ile Ser Tyr Leu Glu Ser His Thr Lys His Asn Asn Ser
530 535 540

CAT TGAAAGTGTC TGACCTATCC TCTAGCAGTT GTCGACAAAT TTCTCCAGTT 1745
His
545

CATGTACAGT AGAAACCGAT GCGTTGCAGT TTCAGAACAT CTTCACTTCT TCAGAGATTA 1805

ACCCTTCGTT GAACATCCAC CAGAAAGGTA GTCACATGTG TAAGTGGGAA AATGAGGTTA 1865

AAAACTATTA TGGCGGCCGA AATGTTCCTT TTTGTTTTCC TCACAAGTGG CCTACGGCAC 1925

TTGATGTTGG AAATACATTT AAATTTGTTG AATTGTTTGA GAACACATGC GTGACGTGTA 1985

ATATTTGCCT ATTGTGATTT TAGCAGTAGT CTTGGCCAGA TTATGCTTTA CGCCTTTAAA 2045

AAAAAAAAAA AAAAAA 2061

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 544 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Leu Ala Leu Thr Ala Ser Ala Ser Ser Ala Ser Ser His Pro Tyr
1 5 10 15

Arg His Ala Ser Ala His Thr Arg Arg Pro Arg Leu Arg Ala Val Leu
20 25 30

Ala Met Ala Gly Ser Asp Asp Pro Arg Ala Ala Pro Ala Arg Ser Val
35 40 45

Ala Val Val Gly Ala Gly Val Ser Gly Leu Ala Ala Ala Tyr Arg Leu
50 55 60

Arg Gln Ser Gly Val Asn Val Thr Val Phe Glu Ala Ala Asp Arg Ala
65 70 75 80

Gly Gly Lys Ile Arg Thr Asn Ser Glu Gly Gly Phe Val Trp Asp Glu
85 90 95

Gly Ala Asn Thr Met Thr Glu Gly Glu Trp Glu Ala Ser Arg Leu Ile
100 105 110

```
Asp Asp Leu Gly Leu Gln Asp Lys Gln Gln Tyr Pro Asn Ser Gln His
        115                 120                 125

Lys Arg Tyr Ile Val Lys Asp Gly Ala Pro Ala Leu Ile Pro Ser Asp
    130                 135                 140

Pro Ile Ser Leu Met Lys Ser Ser Val Leu Ser Thr Lys Ser Lys Ile
145                 150                 155                 160

Ala Leu Phe Phe Glu Pro Phe Leu Tyr Lys Lys Ala Asn Thr Arg Asn
                165                 170                 175

Ser Gly Lys Val Ser Glu Glu His Leu Ser Glu Ser Val Gly Ser Phe
            180                 185                 190

Cys Glu Arg His Phe Gly Arg Glu Val Val Asp Tyr Phe Val Asp Pro
        195                 200                 205

Phe Val Ala Gly Thr Ser Ala Gly Asp Pro Glu Ser Leu Ser Ile Arg
    210                 215                 220

His Ala Phe Pro Ala Leu Trp Asn Leu Glu Arg Lys Tyr Gly Ser Val
225                 230                 235                 240

Ile Val Gly Ala Ile Leu Ser Lys Leu Ala Ala Lys Gly Asp Pro Val
                245                 250                 255

Lys Thr Arg His Asp Ser Ser Gly Lys Arg Arg Asn Arg Arg Val Ser
            260                 265                 270

Phe Ser Phe His Gly Gly Met Gln Ser Leu Ile Asn Ala Leu His Asn
        275                 280                 285

Glu Val Gly Asp Asp Asn Val Lys Leu Gly Thr Glu Val Leu Ser Leu
    290                 295                 300

Ala Cys Thr Phe Asp Gly Val Pro Ala Leu Gly Arg Trp Ser Ile Ser
305                 310                 315                 320

Val Asp Ser Lys Asp Ser Gly Asp Lys Asp Leu Ala Ser Asn Gln Thr
                325                 330                 335

Phe Asp Ala Val Ile Met Thr Ala Pro Leu Ser Asn Val Arg Arg Met
            340                 345                 350

Lys Phe Thr Lys Gly Gly Ala Pro Val Val Leu Asp Phe Leu Pro Lys
        355                 360                 365

Met Asp Tyr Leu Pro Leu Ser Leu Met Val Thr Ala Phe Lys Lys Asp
    370                 375                 380

Asp Val Lys Lys Pro Leu Glu Gly Phe Gly Val Leu Ile Pro Tyr Lys
385                 390                 395                 400

Glu Gln Gln Lys His Gly Leu Lys Thr Leu Gly Thr Leu Phe Ser Ser
                405                 410                 415

Met Met Phe Pro Asp Arg Ala Pro Asp Asp Gln Tyr Leu Tyr Thr Thr
            420                 425                 430

Phe Val Gly Gly Ser His Asn Arg Asp Leu Ala Gly Ala Pro Thr Ser
        435                 440                 445
```

```
Il  Leu Lys Gln Leu Val Thr Ser Asp Leu Lys Lys Leu Leu Gly Val
    450                 455                 460

Glu Gly Gln Pro Thr Phe Val Lys His Val Tyr Trp Gly Asn Ala Phe
465                 470                 475                 480

Pro Leu Tyr Gly His Asp Tyr Ser Ser Val Leu Glu Ala Ile Glu Lys
                485                 490                 495

Met Glu Lys Asn Leu Pro Gly Phe Phe Tyr Ala Gly Asn Ser Lys Asp
            500                 505                 510

Gly Leu Ala Val Gly Ser Val Ile Ala Ser Gly Ser Lys Ala Ala Asp
        515                 520                 525

Leu Ala Ile Ser Tyr Leu Glu Ser His Thr Lys His Asn Asn Ser His
    530                 535                 540
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1811 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..1589
(D) OTHER INFORMATION: /product= "wheat protox-1 cDNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
GC GCA ACA ATG GCC ACC GCC ACC GTC GCG GCC GCG TCG CCG CTC CGC          47
   Ala Thr Met Ala Thr Ala Thr Val Ala Ala Ala Ser Pro Leu Arg
     1               5                  10                  15

GGC AGG GTC ACC GGG CGC CCA CAC CGC GTC CGC CCG CGT TGC GCT ACC         95
Gly Arg Val Thr Gly Arg Pro His Arg Val Arg Pro Arg Cys Ala Thr
                20                  25                  30

GCG AGC AGC GCG ACC GAG ACT CCG GCG GCG CCC GGC GTG CGG CTG TCC        143
Ala Ser Ser Ala Thr Glu Thr Pro Ala Ala Pro Gly Val Arg Leu Ser
            35                  40                  45

GCG GAA TGC GTC ATT GTG GGC GCC GGC ATC AGC GGC CTC TGC ACC GCG        191
Ala Glu Cys Val Ile Val Gly Ala Gly Ile Ser Gly Leu Cys Thr Ala
        50                  55                  60

CAG GCG CTG GCC ACC CGA TAC GGC GTC AGC GAC CTG CTC GTC ACG GAG        239
Gln Ala Leu Ala Thr Arg Tyr Gly Val Ser Asp Leu Leu Val Thr Glu
    65                  70                  75

GCC CGC GAC CGC CCG GGC GGC AAC ATC ACC ACC GTC GAG CGT CCC GAC        287
```

```
Ala Arg Asp Arg Pro Gly Gly Asn Il  Thr Thr Val Glu Arg Pro Asp
80                  85                  90                  95

GAG GGG TAC CTG TGG GAG GAG GGA CCC AAC AGC TTC CAG CCC TCC GAC         335
Glu Gly Tyr Leu Trp Glu Glu Gly Pro Asn Ser Phe Gln Pro Ser Asp
                100                 105                 110

CCG GTC CTC ACC ATG GCC GTG GAC AGC GGG CTC AAG GAT GAC TTG GTG         383
Pro Val Leu Thr Met Ala Val Asp Ser Gly Leu Lys Asp Asp Leu Val
            115                 120                 125

TTC GGG GAC CCC AAC GCG CCC CGG TTC GTG CTG TGG GAG GGG AAG CTG         431
Phe Gly Asp Pro Asn Ala Pro Arg Phe Val Leu Trp Glu Gly Lys Leu
        130                 135                 140

AGG CCG GTG CCG TCG AAG CCA GGC GAC CTG CCT TTC TTC AGC CTC ATG         479
Arg Pro Val Pro Ser Lys Pro Gly Asp Leu Pro Phe Phe Ser Leu Met
    145                 150                 155

AGT ATC CCT GGG AAG CTC AGG GCC GGC CTT GGC GCG CTC GGC ATT CGC         527
Ser Ile Pro Gly Lys Leu Arg Ala Gly Leu Gly Ala Leu Gly Ile Arg
160                 165                 170                 175

CCA CCT CCT CCA GGG CGC GAG GAG TCG GTG GAG GAG TTT GTG CGC CGC         575
Pro Pro Pro Pro Gly Arg Glu Glu Ser Val Glu Glu Phe Val Arg Arg
                180                 185                 190

AAC CTC GGT GCC GAG GTC TTT GAG CGC CTC ATC GAG CCT TTC TGC TCA         623
Asn Leu Gly Ala Glu Val Phe Glu Arg Leu Ile Glu Pro Phe Cys Ser
            195                 200                 205

GGT GTA TAT GCT GGT GAT CCT TCG AAG CTT AGT ATG AAG GCT GCA TTT         671
Gly Val Tyr Ala Gly Asp Pro Ser Lys Leu Ser Met Lys Ala Ala Phe
        210                 215                 220

GGG AAG GTC TGG AGG TTG GAG GAG ATT GGA GGT AGT ATT ATT GGT GGA         719
Gly Lys Val Trp Arg Leu Glu Glu Ile Gly Gly Ser Ile Ile Gly Gly
    225                 230                 235

ACC ATC AAG GCG ATT CAG GAT AAA GGG AAG AAC CCC AAA CCG CCA AGG         767
Thr Ile Lys Ala Ile Gln Asp Lys Gly Lys Asn Pro Lys Pro Pro Arg
240                 245                 250                 255

GAT CCC CGA CTT CCG GCA CCA AAG GGA CAG ACG GTG GCA TCT TTC AGG         815
Asp Pro Arg Leu Pro Ala Pro Lys Gly Gln Thr Val Ala Ser Phe Arg
                260                 265                 270

AAG GGT CTA GCC ATG CTC CCG AAT GCC ATC GCA TCT AGG CTG GGT AGT         863
Lys Gly Leu Ala Met Leu Pro Asn Ala Ile Ala Ser Arg Leu Gly Ser
            275                 280                 285

AAA GTC AAG CTG TCA TGG AAG CTT ACG AGC ATT ACA AAG GCG GAC AAC         911
Lys Val Lys Leu Ser Trp Lys Leu Thr Ser Ile Thr Lys Ala Asp Asn
        290                 295                 300

CAA GGA TAT GTA TTA GGT TAT GAA ACA CCA GAA GGA CTT GTT TCA GTG         959
Gln Gly Tyr Val Leu Gly Tyr Glu Thr Pro Glu Gly Leu Val Ser Val
    305                 310                 315

CAG GCT AAA AGT GTT ATC ATG ACC ATC CCG TCA TAT GTT GCT AGT GAT        1007
Gln Ala Lys Ser Val Ile Met Thr Ile Pro Ser Tyr Val Ala Ser Asp
320                 325                 330                 335
```

```
ATC TTG CGC CCA CTT TCA ATT GAT GCA GCA GAT GCA CTC TCA AAA TTC      1055
Il  Leu Arg Pro Leu Ser Ile Asp Ala Ala Asp Ala Leu Ser Lys Phe
                340                 345                 350

TAT TAT CCG CCA GTT GCT GCT GTA ACT GTT TCA TAT CCA AAA GAA GCT      1103
Tyr Tyr Pro Pro Val Ala Ala Val Thr Val Ser Tyr Pro Lys Glu Ala
            355                 360                 365

ATT AGA AAA GAA TGC TTA ATT GAT GGG GAG CTC CAG GGT TTC GGC CAG      1151
Ile Arg Lys Glu Cys Leu Ile Asp Gly Glu Leu Gln Gly Phe Gly Gln
        370                 375                 380

TTG CAT CCA CGT AGC CAA GGA GTC GAG ACT TTA GGG ACA ATA TAT AGC      1199
Leu His Pro Arg Ser Gln Gly Val Glu Thr Leu Gly Thr Ile Tyr Ser
    385                 390                 395

TCT TCT CTC TTT CCT AAT CGT GCT CCT GCT GGA AGA GTG TTA CTT CTG      1247
Ser Ser Leu Phe Pro Asn Arg Ala Pro Ala Gly Arg Val Leu Leu Leu
400                 405                 410                 415

AAC TAT ATC GGG GGT TCT ACA AAT ACA GGG ATC GTC TCC AAG ACT GAG      1295
Asn Tyr Ile Gly Gly Ser Thr Asn Thr Gly Ile Val Ser Lys Thr Glu
                420                 425                 430

AGT GAC TTA GTA GGA GCC GTT GAC CGT GAC CTC AGA AAA ATG TTG ATA      1343
Ser Asp Leu Val Gly Ala Val Asp Arg Asp Leu Arg Lys Met Leu Ile
            435                 440                 445

AAC CCT AGA GCA GCA GAC CCT TTA GCA TTA GGG GTT CGA GTG TGG CCA      1391
Asn Pro Arg Ala Ala Asp Pro Leu Ala Leu Gly Val Arg Val Trp Pro
        450                 455                 460

CAA GCA ATA CCA CAG TTT TTG ATT GGG CAC CTT GAT CGC CTT GCT GCT      1439
Gln Ala Ile Pro Gln Phe Leu Ile Gly His Leu Asp Arg Leu Ala Ala
    465                 470                 475

GCA AAA TCT GCA CTG GGC CAA GGC GGC TAC GAC GGG TTG TTC CTA GGA      1487
Ala Lys Ser Ala Leu Gly Gln Gly Gly Tyr Asp Gly Leu Phe Leu Gly
480                 485                 490                 495

GGA AAC TAC GTC GCA GGA GTT GCC TTG GGC CGA TGC ATC GAG GGT GCG      1535
Gly Asn Tyr Val Ala Gly Val Ala Leu Gly Arg Cys Ile Glu Gly Ala
                500                 505                 510

TAC GAG AGT GCC TCA CAA GTA TCT GAC TTC TTG ACC AAG TAT GCC TAC      1583
Tyr Glu Ser Ala Ser Gln Val Ser Asp Phe Leu Thr Lys Tyr Ala Tyr
            515                 520                 525

AAG TGA TGGAAGTAGT GCATCTCTTC ATTTTGTTGC ATATACGAGG TGAGGCTAGG        1639
Lys

ATCGGTAAAA CATCATGAGA TTCTGTAGTG TTTCTTTAAT TGAAAAAACA AATTTTAGTG    1699

ATGCAATATG TGCTCTTTCC TGTAGTTCGA GCATGTACAT CGGTATGGGA TAAAGTAGAA    1759

TAAGCTATTC TGCAAAAGCA GTGATTTTTT TTGAAAAAAA AAAAAAAAAA AA            1811
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 528 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Ala Thr Met Ala Thr Ala Thr Val Ala Ala Ala Ser Pro Leu Arg Gly
1               5                   10                  15

Arg Val Thr Gly Arg Pro His Arg Val Arg Pro Arg Cys Ala Thr Ala
            20                  25                  30

Ser Ser Ala Thr Glu Thr Pro Ala Ala Pro Gly Val Arg Leu Ser Ala
        35                  40                  45

Glu Cys Val Ile Val Gly Ala Gly Ile Ser Gly Leu Cys Thr Ala Gln
    50                  55                  60

Ala Leu Ala Thr Arg Tyr Gly Val Ser Asp Leu Leu Val Thr Glu Ala
65                  70                  75                  80

Arg Asp Arg Pro Gly Gly Asn Ile Thr Thr Val Glu Arg Pro Asp Glu
                85                  90                  95

Gly Tyr Leu Trp Glu Glu Gly Pro Asn Ser Phe Gln Pro Ser Asp Pro
            100                 105                 110

Val Leu Thr Met Ala Val Asp Ser Gly Leu Lys Asp Asp Leu Val Phe
        115                 120                 125

Gly Asp Pro Asn Ala Pro Arg Phe Val Leu Trp Glu Gly Lys Leu Arg
    130                 135                 140

Pro Val Pro Ser Lys Pro Gly Asp Leu Pro Phe Phe Ser Leu Met Ser
145                 150                 155                 160

Ile Pro Gly Lys Leu Arg Ala Gly Leu Gly Ala Leu Gly Ile Arg Pro
                165                 170                 175

Pro Pro Pro Gly Arg Glu Glu Ser Val Glu Glu Phe Val Arg Arg Asn
            180                 185                 190

Leu Gly Ala Glu Val Phe Glu Arg Leu Ile Glu Pro Phe Cys Ser Gly
        195                 200                 205

Val Tyr Ala Gly Asp Pro Ser Lys Leu Ser Met Lys Ala Ala Phe Gly
    210                 215                 220

Lys Val Trp Arg Leu Glu Glu Ile Gly Gly Ser Ile Ile Gly Gly Thr
225                 230                 235                 240

Ile Lys Ala Ile Gln Asp Lys Gly Lys Asn Pro Lys Pro Pro Arg Asp
                245                 250                 255

Pro Arg Leu Pro Ala Pro Lys Gly Gln Thr Val Ala Ser Phe Arg Lys
            260                 265                 270

Gly Leu Ala Met Leu Pro Asn Ala Ile Ala Ser Arg Leu Gly Ser Lys
        275                 280                 285

Val Lys Leu Ser Trp Lys Leu Thr Ser Ile Thr Lys Ala Asp Asn Gln
    290                 295                 300
```

```
Gly Tyr Val Leu Gly Tyr Glu Thr Pro Glu Gly Leu Val Ser Val Gln
305                 310                 315                 320

Al  Lys Ser Val Ile Met Thr Ile Pro Ser Tyr Val Ala Ser Asp Ile
                325                 330                 335

Leu Arg Pro Leu Ser Ile Asp Ala Ala Asp Ala Leu Ser Lys Phe Tyr
            340                 345                 350

Tyr Pro Pro Val Ala Ala Val Thr Val Ser Tyr Pro Lys Glu Ala Ile
        355                 360                 365

Arg Lys Glu Cys Leu Ile Asp Gly Glu Leu Gln Gly Phe Gly Gln Leu
    370                 375                 380

His Pro Arg Ser Gln Gly Val Glu Thr Leu Gly Thr Ile Tyr Ser Ser
385                 390                 395                 400

Ser Leu Phe Pro Asn Arg Ala Pro Ala Gly Arg Val Leu Leu Leu Asn
                405                 410                 415

Tyr Ile Gly Gly Ser Thr Asn Thr Gly Ile Val Ser Lys Thr Glu Ser
            420                 425                 430

Asp Leu Val Gly Ala Val Asp Arg Asp Leu Arg Lys Met Leu Ile Asn
        435                 440                 445

Pro Arg Ala Ala Asp Pro Leu Ala Leu Gly Val Arg Val Trp Pro Gln
    450                 455                 460

Ala Ile Pro Gln Phe Leu Ile Gly His Leu Asp Arg Leu Ala Ala Ala
465                 470                 475                 480

Lys Ser Ala Leu Gly Gln Gly Gly Tyr Asp Gly Leu Phe Leu Gly Gly
                485                 490                 495

Asn Tyr Val Ala Gly Val Ala Leu Gly Arg Cys Ile Glu Gly Ala Tyr
            500                 505                 510

Glu Ser Ala Ser Gln Val Ser Asp Phe Leu Thr Lys Tyr Ala Tyr Lys
        515                 520                 525
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1847 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:
- (A) NAME/KEY: CDS
- (B) LOCATION: 55..1683
- (D) OTHER INFORMATION: /product= "soybean protox-1 cDNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
TTTTAGCACA GTGTTGAAGA TAACGAACGA ATAGTCCCAT TACTGTAACC AACC ATG      57
                                                            Met
                                                              1

GTT TCC GTC TTC AAC GAG ATC CTA TTC CCG CCG AAC CAA ACC CTT CTT     105
Val Ser Val Phe Asn Glu Ile Leu Phe Pro Pro Asn Gln Thr Leu Leu
                  5                  10                  15

CGC CCC TCC CTC CAT TCC CCA ACC TCT TTC TTC ACC TCT CCC ACT CGA     153
Arg Pro Ser Leu His Ser Pro Thr Ser Phe Phe Thr Ser Pro Thr Arg
         20                  25                  30

AAA TTC CCT CGC TCT CGC CCT AAC CCT ATT CTA CGC TGC TCC ATT GCG     201
Lys Phe Pro Arg Ser Arg Pro Asn Pro Ile Leu Arg Cys Ser Ile Ala
     35                  40                  45

GAG GAA TCC ACC GCG TCT CCG CCC AAA ACC AGA GAC TCC GCC CCC GTG     249
Glu Glu Ser Thr Ala Ser Pro Pro Lys Thr Arg Asp Ser Ala Pro Val
 50                  55                  60                  65

GAC TGC GTC GTC GTC GGC GGA GGC GTC AGC GGC CTC TGC ATC GCC CAG     297
Asp Cys Val Val Val Gly Gly Gly Val Ser Gly Leu Cys Ile Ala Gln
                 70                  75                  80

GCC CTC GCC ACC AAA CAC GCC AAT GCC AAC GTC GTC GTC ACG GAG GCC     345
Ala Leu Ala Thr Lys His Ala Asn Ala Asn Val Val Val Thr Glu Ala
             85                  90                  95

CGA GAC CGC GTC GGC GGC AAC ATC ACC ACG ATG GAG AGG GAC GGA TAC     393
Arg Asp Arg Val Gly Gly Asn Ile Thr Thr Met Glu Arg Asp Gly Tyr
        100                 105                 110

CTC TGG GAA GAA GGC CCC AAC AGC TTC CAG CCT TCT GAT CCA ATG CTC     441
Leu Trp Glu Glu Gly Pro Asn Ser Phe Gln Pro Ser Asp Pro Met Leu
    115                 120                 125

ACC ATG GTG GTG GAC AGT GGT TTA AAG GAT GAG CTT GTT TTG GGG GAT     489
Thr Met Val Val Asp Ser Gly Leu Lys Asp Glu Leu Val Leu Gly Asp
130                 135                 140                 145

CCT GAT GCA CCT CGG TTT GTG TTG TGG AAC AGG AAG TTG AGG CCG GTG     537
Pro Asp Ala Pro Arg Phe Val Leu Trp Asn Arg Lys Leu Arg Pro Val
                150                 155                 160

CCC GGG AAG CTG ACT GAT TTG CCT TTC TTT GAC TTG ATG AGC ATT GGT     585
Pro Gly Lys Leu Thr Asp Leu Pro Phe Phe Asp Leu Met Ser Ile Gly
            165                 170                 175

GGC AAA ATC AGG GCT GGC TTT GGT GCG CTT GGA ATT CGG CCT CCT CCT     633
Gly Lys Ile Arg Ala Gly Phe Gly Ala Leu Gly Ile Arg Pro Pro Pro
        180                 185                 190

CCA GGT CAT GAG GAA TCG GTT GAA GAG TTT GTT CGT CGG AAC CTT GGT     681
Pro Gly His Glu Glu Ser Val Glu Glu Phe Val Arg Arg Asn Leu Gly
    195                 200                 205

GAT GAG GTT TTT GAA CGG TTG ATA GAG CCT TTT TGT TCA GGG GTC TAT     729
Asp Glu Val Phe Glu Arg Leu Ile Glu Pro Phe Cys Ser Gly Val Tyr
210                 215                 220                 225

GCA GGC GAT CCT TCA AAA TTA AGT ATG AAA GCA GCA TTC GGG AAA GTT     777
```

```
Ala Gly Asp Pro Ser Lys Leu Ser Met Lys Ala Ala Phe Gly Lys Val
                230                 235                 240

TGG AAG CTG GAA AAA AAT GGT GGT AGC ATT ATT GGT GGA ACT TTC AAA      825
Trp Lys Leu Glu Lys Asn Gly Gly Ser Ile Ile Gly Gly Thr Phe Lys
            245                 250                 255

GCA ATA CAA GAG AGA AAT GGA GCT TCA AAA CCA CCT CGA GAT CCG CGT      873
Ala Ile Gln Glu Arg Asn Gly Ala Ser Lys Pro Pro Arg Asp Pro Arg
        260                 265                 270

CTG CCA AAA CCA AAA GGT CAG ACT GTT GGA TCT TTC CGG AAG GGA CTT      921
Leu Pro Lys Pro Lys Gly Gln Thr Val Gly Ser Phe Arg Lys Gly Leu
    275                 280                 285

ACC ATG TTG CCT GAT GCA ATT TCT GCC AGA CTA GGC AAC AAA GTA AAG      969
Thr Met Leu Pro Asp Ala Ile Ser Ala Arg Leu Gly Asn Lys Val Lys
290                 295                 300                 305

TTA TCT TGG AAG CTT TCA AGT ATT AGT AAA CTG GAT AGT GGA GAG TAC     1017
Leu Ser Trp Lys Leu Ser Ser Ile Ser Lys Leu Asp Ser Gly Glu Tyr
                310                 315                 320

AGT TTG ACA TAT GAA ACA CCA GAA GGA GTG GTT TCT TTG CAG TGC AAA     1065
Ser Leu Thr Tyr Glu Thr Pro Glu Gly Val Val Ser Leu Gln Cys Lys
            325                 330                 335

ACT GTT GTC CTG ACC ATT CCT TCC TAT GTT GCT AGT ACA TTG CTG CGT     1113
Thr Val Val Leu Thr Ile Pro Ser Tyr Val Ala Ser Thr Leu Leu Arg
        340                 345                 350

CCT CTG TCT GCT GCT GCT GCA GAT GCA CTT TCA AAG TTT TAT TAC CCT     1161
Pro Leu Ser Ala Ala Ala Ala Asp Ala Leu Ser Lys Phe Tyr Tyr Pro
    355                 360                 365

CCA GTT GCT GCA GTT TCC ATA TCC TAT CCA AAA GAA GCT ATT AGA TCA     1209
Pro Val Ala Ala Val Ser Ile Ser Tyr Pro Lys Glu Ala Ile Arg Ser
370                 375                 380                 385

GAA TGC TTG ATA GAT GGT GAG TTG AAG GGG TTT GGT CAA TTG CAT CCA     1257
Glu Cys Leu Ile Asp Gly Glu Leu Lys Gly Phe Gly Gln Leu His Pro
                390                 395                 400

CGT AGC CAA GGA GTG GAA ACA TTA GGA ACT ATA TAC AGC TCA TCA CTA     1305
Arg Ser Gln Gly Val Glu Thr Leu Gly Thr Ile Tyr Ser Ser Ser Leu
            405                 410                 415

TTC CCC AAC CGA GCA CCA CCT GGA AGG GTT CTA CTC TTG AAT TAC ATT     1353
Phe Pro Asn Arg Ala Pro Pro Gly Arg Val Leu Leu Leu Asn Tyr Ile
        420                 425                 430

GGA GGA GCA ACT AAT ACT GGA ATT TTA TCG AAG ACG GAC AGT GAA CTT     1401
Gly Gly Ala Thr Asn Thr Gly Ile Leu Ser Lys Thr Asp Ser Glu Leu
    435                 440                 445

GTG GAA ACA GTT GAT CGA GAT TTG AGG AAA ATC CTT ATA AAC CCA AAT     1449
Val Glu Thr Val Asp Arg Asp Leu Arg Lys Ile Leu Ile Asn Pro Asn
450                 455                 460                 465

GCC CAG GAT CCA TTT GTA GTG GGG GTG AGA CTG TGG CCT CAA GCT ATT     1497
Ala Gln Asp Pro Phe Val Val Gly Val Arg Leu Trp Pro Gln Ala Ile
                470                 475                 480
```

```
CCA CAG TTC TTA GTT GGC CAT CTT GAT CTT CTA GAT GTT GCT AAA GCT      1545
Pro Gln Phe Leu Val Gly His Leu Asp Leu Leu Asp Val Ala Lys Ala
            485                 490                 495

TCT ATC AGA AAT ACT GGG TTT GAA GGG CTC TTC CTT GGG GGT AAT TAT      1593
Ser Ile Arg Asn Thr Gly Phe Glu Gly Leu Phe Leu Gly Gly Asn Tyr
        500                 505                 510

GTG TCT GGT GTT GCC TTG GGA CGA TGC GTT GAG GGA GCC TAT GAG GTA      1641
Val Ser Gly Val Ala Leu Gly Arg Cys Val Glu Gly Ala Tyr Glu Val
    515                 520                 525

GCA GCT GAA GTA AAC GAT TTT CTC ACA AAT AGA GTG TAC AAA              1683
Ala Ala Glu Val Asn Asp Phe Leu Thr Asn Arg Val Tyr Lys
530                 535                 540

TAGTAGCAGT TTTTGTTTTT GTGGTGGAAT GGGTGATGGG ACTCTCCTGT TCCATTGAAT    1743

TATAATAATG TGAAAGTTTC TCAAATTCGT TCGATAGGTT TTTGGCGGCT TCTATTGCTG    1803

ATAATGTAAA ATCCTCTTTA AGTTTGAAAA AAAAAAAAAA AAAA                     1847
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 543 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID 12:

```
Met Val Ser Val Phe Asn Glu Ile Leu Phe Pro Pro Asn Gln Thr Leu
1               5                   10                  15

Leu Arg Pro Ser Leu His Ser Pro Thr Ser Phe Phe Thr Ser Pro Thr
            20                  25                  30

Arg Lys Phe Pro Arg Ser Arg Pro Asn Pro Ile Leu Arg Cys Ser Ile
        35                  40                  45

Ala Glu Glu Ser Thr Ala Ser Pro Pro Lys Thr Arg Asp Ser Ala Pro
    50                  55                  60

Val Asp Cys Val Val Val Gly Gly Gly Val Ser Gly Leu Cys Ile Ala
65                  70                  75                  80

Gln Ala Leu Ala Thr Lys His Ala Asn Ala Asn Val Val Val Thr Glu
                85                  90                  95

Ala Arg Asp Arg Val Gly Gly Asn Ile Thr Thr Met Glu Arg Asp Gly
            100                 105                 110

Tyr Leu Trp Glu Glu Gly Pro Asn Ser Phe Gln Pro Ser Asp Pro Met
        115                 120                 125

Leu Thr Met Val Val Asp Ser Gly Leu Lys Asp Glu Leu Val Leu Gly
    130                 135                 140

Asp Pro Asp Ala Pro Arg Phe Val Leu Trp Asn Arg Lys Leu Arg Pro
145                 150                 155                 160
```

```
Val Pro Gly Lys Leu Thr Asp Leu Pro Phe Phe Asp Leu Met Ser Ile
                165                 170                 175

Gly Gly Lys Ile Arg Ala Gly Phe Gly Ala Leu Gly Ile Arg Pro Pro
            180                 185                 190

Pro Pro Gly His Glu Glu Ser Val Glu Glu Phe Val Arg Arg Asn Leu
        195                 200                 205

Gly Asp Glu Val Phe Glu Arg Leu Ile Glu Pro Phe Cys Ser Gly Val
    210                 215                 220

Tyr Ala Gly Asp Pro Ser Lys Leu Ser Met Lys Ala Ala Phe Gly Lys
225                 230                 235                 240

Val Trp Lys Leu Glu Lys Asn Gly Gly Ser Ile Ile Gly Gly Thr Phe
                245                 250                 255

Lys Ala Ile Gln Glu Arg Asn Gly Ala Ser Lys Pro Pro Arg Asp Pro
            260                 265                 270

Arg Leu Pro Lys Pro Lys Gly Gln Thr Val Gly Ser Phe Arg Lys Gly
        275                 280                 285

Leu Thr Met Leu Pro Asp Ala Ile Ser Ala Arg Leu Gly Asn Lys Val
    290                 295                 300

Lys Leu Ser Trp Lys Leu Ser Ser Ile Ser Lys Leu Asp Ser Gly Glu
305                 310                 315                 320

Tyr Ser Leu Thr Tyr Glu Thr Pro Glu Gly Val Val Ser Leu Gln Cys
                325                 330                 335

Lys Thr Val Val Leu Thr Ile Pro Ser Tyr Val Ala Ser Thr Leu Leu
            340                 345                 350

Arg Pro Leu Ser Ala Ala Ala Ala Asp Ala Leu Ser Lys Phe Tyr Tyr
        355                 360                 365

Pro Pro Val Ala Ala Val Ser Ile Ser Tyr Pro Lys Glu Ala Ile Arg
    370                 375                 380

Ser Glu Cys Leu Ile Asp Gly Glu Leu Lys Gly Phe Gly Gln Leu His
385                 390                 395                 400

Pro Arg Ser Gln Gly Val Glu Thr Leu Gly Thr Ile Tyr Ser Ser Ser
                405                 410                 415

Leu Phe Pro Asn Arg Ala Pro Pro Gly Arg Val Leu Leu Leu Asn Tyr
            420                 425                 430

Ile Gly Gly Ala Thr Asn Thr Gly Ile Leu Ser Lys Thr Asp Ser Glu
        435                 440                 445

Leu Val Glu Thr Val Asp Arg Asn Leu Arg Lys Ile Leu Ile Asn Pro
    450                 455                 460

Asn Ala Gln Asp Pro Phe Val Val Gly Val Arg Leu Trp Pro Gln Ala
465                 470                 475                 480

Ile Pro Gln Phe Leu Val Gly His Leu Asp Leu Leu Asp Val Ala Lys
                485                 490                 495
```

```
Ala Ser Ile Arg Asn Thr Gly Phe Glu Gly Leu Phe Leu Gly Gly Asn
            500                 505                 510

Tyr Val Ser Gly Val Ala Leu Gly Arg Cys Val Glu Gly Ala Tyr Glu
        515                 520                 525

Val Ala Ala Glu Val Asn Asp Phe Leu Thr Asn Arg Val Tyr Lys
    530                 535                 540
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 583 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(ix) FEATURE:
(A) NAME/KEY: promoter
(B) LOCATION: 1..583
(D) OTHER INFORMATION: /function= "arabidopsis protox-1 promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
GAATTCCGAT CGAATTATAT AATTATCATA AATTTGAATA AGCATGTTGC CTTTTATTAA     60
AGAGGTTTAA TAAAGTTTGG TAATAATGGA CTTTGACTTC AAACTCGATT CTCATGTAAT    120
TAATTAATAT TTACATCAAA ATTTGGTCAC TAATATTACC AAATTAATAT ACTAAAATGT    180
TAATTCGCAA ATAAAACACT AATTCCAAAT AAAGGGTCAT TATGATAAAC ACGTATTGAA    240
CTTGATAAAG CAAAGCAAAA ATAATGGGTT TCAAGGTTTG GGTTATATAT GACAAAAAAA    300
AAAAAAGGTT TGGTTATATA TCTATTGGGC CTATAACCAT GTTATACAAA TTTGGGCCTA    360
ACTAAAATAA TAAAATAAAC GTAATGGTCC TTTTTATATT TGGGTCAAAC CCAACTCTAA    420
ACCCAAACCA AAGAAAAAGT ATACGGTACG GTACACAGAC TTATGGTGTG TGTGATTGCA    480
GGTGAATATT TCTCGTCGTC TTCTCCTTTC TTCTGAAGAA GATTACCCAA TCTGAAAAAA    540
ACCAAGAAGC TGACAAAATT CCGAATTCTC TGGGATTTCC ATG                      583
```

The invention as described herein is contemplated to include the following enumerated embodiments:

1. A recombinant DNA molecule comprising a plant protoporphyrinogen oxidase (protox) promoter or a functionally equivalent derivative thereof.

2. A chimeric gene comprising a plant protox promoter operably linked to a heterologous DNA coding sequence.

3. The chimeric gene of claim 2 wherein said plant protox promoter is from a protox-1 gene.

4. The chimeric gene of claim 2 wherein said plant protox promoter is from a protox-2 gene.

5. The chimeric gene of claim 2 wherein said protox promoter is from a plant selected from the group consisting of *Arabidopsis*, soybean, cotton, tobacco, sugar beet, oilseed rape, maize, wheat, sorghum, rye, oats, turf grass and rice.

6. The chimeric gene of claim 5 wherein said promoter is from an *Arabidopsis* plant.

7. The chimeric gene of claim 6 wherein said promoter is at least 300 nucleotides in length.

8. The chimeric gene of claim 7 wherein said promoter is at least 500 nucleotides in length.

9. The chimeric gene of claim 8 wherein said promoter has the sequence set forth in SEQ ID No. 13.

10. The chimeric gene of claim 2 wherein said heterologous coding sequence encodes a modified, herbicide-resistant form of a plant enzyme.

11. The chimeric gene of claim 10 wherein said plant enzyme is selected from the group consisting of imidazoleglycerol phosphate dehydratase (IGPD), EPSP synthase, glutamine synthetase (GS), acetyl coenzyme A carboxylase, acetolactate synthase, and protoporphyrinogen oxidase (protox).

12. The chimeric gene of claim 11 wherein said plant enzyme is protox.

13. A recombinant DNA vector comprising the recombinant DNA molecule of claim 1.

14. Plant tissue comprising the chimeric gene of claim 2.

15. A plant comprising the chimeric gene of claim 2.

16. The plant of claim 15 wherein said plant is selected from the group consisting of *Arabidopsis*, soybean, cotton, tobacco, sugar beet, oilseed rape, maize, wheat, sorghum, rye, oats, turf grass and rice.

## ABSTRACT OF DISCLOSURE

Promoters naturally associated with plant protoporphyrinogen oxidase (protox) coding sequences, and derivatives thereof, are provided. These promoters can be used to control the expression of an operably linked heterologous coding sequence in a plant cell. . These promoters are particularly useful for expressing modified forms of herbicide target enzymes, particularly modified forms of protox, to achieve tolerance to herbicides which inhibit the corresponding unmodified enzymes. Recombinant DNA molecules and chimeric genes comprising these promoters are provided, as well as plant tissue and plants containing such chimeric genes.